APPLY THE LINDDUN FRAMEWORK FOR PRIVACY REQUIREMENT ANALYSIS

Pengfei Lu

University of Tampere
School of Information Sciences
Master Degree Program in Software Development
M. Sc. thesis
Supervisor: Zheying Zhang
March, 2017

University of Tampere
Faculty of Natural Sciences
Computer Science / Software development
Pengfei Lu: Apply the LINDDUN framework for privacy requirement analysis
M.Sc. thesis, 56 pages, 11 index and appendix pages
March, 2017

LINDDUN is a framework to identify privacy threats and elicit privacy requirements from a system. It has complete procedures and strong support on privacy requirements analysis. This research tries to figure out how practically we can apply the LINDDUN methodology in privacy requirements analysis. This thesis studies LINDDUN in a case project name Rin-Tin-Tinder for privacy threats and privacy requirements analysis. The analysis results are compared with the privacy requirement elicited by the project team in a workshop session. The analysis result is verified through a comparison with the Microsoft privacy guideline.

The discussions and analysis on comparison implies strengths and weakness of the LINDDUN methodology. Compared to workshop, the LINDDUN methodology lead the analyst to identify more privacy threats and get more privacy requirements, and makes analyzing process more predictable. Meanwhile, the LINDDUN methodology has a blind spot on users' unintentional false instructions. The thesis discussed possible directions to improve LINDDUN and summarize a guide rules on assumption making, which is an important procedure in LINDDUN. These findings will be helpful for LINDDUN's further improvement.

Key words and terms: LINDDUN, privacy threat, privacy requirement, improvement.

Contents

1. Introduction

1.1. Background

Privacy has been a hotspot for a long period, especially in software fields. In order to make good use of the Internet and computer technology, an increasing number of software products are created and produced. At the meanwhile, privacy becomes a problem to both software development groups and their users. People's explorations on protecting privacy never stop. In 2002, Microsoft made STRIDE[1] threat modeling framework to improve information security. Microsoft privacy guidelines were proposed in 2008 to provide people advices on dealing with privacy issues in software development [Microsoft, 2008]. Building a sufficient privacy policy is also a feasible solution to privacy problems [Dennedy et al., 2014], which has been used in most websites and web applications. Besides, modeling privacy threats is another direction to protect privacy, which is studied by learners and researchers [Beckers, et al., 2014]. This thesis will focus on one privacy threats modeling and privacy requirements analysis methodology: the LINDDUN methodology.

The LINDDUN methodology is one of many methods to model privacy threats and analyze privacy requirements in software development. It contains complete concept framework about privacy threat and detailed procedures to elicit privacy requirements. The name of LINDDUN is an abbreviation of seven privacy threat categories: **L**inkability, **I**dentifiability, **N**on-repudiation, **D**etectability, information **D**isclosure, content **U**nawareness and policy/consent **N**on-compliance. The LINDDUN methodology was built by Mina Deng, et al. [Deng et al., 2010], and developed by Kim Wuyts, et al. [Wuyts, 2015]. The LINDDUN methodology is available and live update online. The version of LINDDUN which is discussed in this thesis is before May, 2016.

1.2. Research objectives

The research objective of this research is to figure out how practically we can apply the LINDDUN methodology in privacy requirements analysis. The LINDDUN methodology is an approach to analyze a system and elicit threats and requirements from this system. The author of LINDDUN introduced a Social network 2.0 sample to explain the LINDDUN methodology in her thesis. Kim Wuyts and her research group has also descript two experimental cases, and came up with suggestions to improve the LINDDUN methodology. However, is there any points or procedures unclear to readers and analysts? Does the LINDDUN methodology have any weakness in privacy

[1] STRIDE is a method to deal with computer security threats. It is developed by Microsoft.

requirement analysis? In order to answer these questions, the process of the LINDDUN methodology shall be studied, analyzed and discussed. The final goal of this thesis is to point out probable weakness and put forward constructive suggestions to improve the LINDDUN methodology.

1.3. Research methods

In order to solve the research questions, a case study is used. Case study is a research method to study a new objective. The objective can be a method, a concept or a theory. Case study is chosen because that the LINDDUN methodology contains a concept framework and a complete method at the same time. It is important to figure out what LINDDUN does and how it works. A social network system, named Rin-Tin-Tinder (RTT), is the study case. The author acts as an analyst to apply the LINDDUN methodology to the RTT system. After analysis, a privacy threat list and a privacy requirement list will be output and discussed for further findings.

Additionally, an interview is made to collect data from development group of the RTT. Before the case system is analyzed, a workshop has been made by three members of the RTT development group to find out privacy threats for the RTT system. The workshop produced a privacy threat list as a result. This result might be valuable to research questions.

1.4. Thesis structure

The thesis consists of three parts, and they are literature review, case study and final conclusion. Chapter 2 contains generic concepts related to privacy issues. Chapter 3 presents an overview of study target: LINDDUN, including concept framework, method procedures, and other necessary methods used in this thesis. In the second part, a case study is introduced and proceed. It produces some result with the help of LINDDUN. Then, in part 3, new findings from part 2 are discussed. Some advices are produced to improve LINDDUN. Finally, in the last chapter, there shall be a conclusion of the whole thesis, and a few words about what further work could be done based on this thesis.

2. Data privacy

Defining the term privacy is difficult in the information science field. The term can be described from different perspectives, and it is difficult to tell which perspective is complete. Legal and policy scholar Alan F. Westin asserted that "no definition … is possible, because privacy issues are fundamentally matters of values, interests and power" [Alan F.W., 1967]. The process of understanding a concept is to put it into a specific application domain. In this chapter, a set of privacy related concepts in software development domain will be introduced and discussed.

2.1. The right to Privacy

When people talk about privacy, in most situations they mean the right to privacy. Warren and Brandeis [1890] articulated the definition and importance of the right to privacy already as early as 1890. By reviewing *"The Right to Privacy"*, the right to privacy is might related to personal rights. They summarized 6 general rules of the right to privacy [p.214–p.218] and 2 remedies for an invasion of the right to privacy [p.219], which makes a big progress on privacy concept. The six rules are: 1.The right to privacy does not prohibit any publication of matter which is of public or general interest; 2.The right to privacy does not prohibit the communication of any matter, though in its nature private, when the publication is made under circumstances which would render it a privileged communication according to the law of slander and libel; 3. The law would probably not grant any redress for the invasion of privacy by oral publication in the absence of special damage; 4. The right to privacy ceases upon the publication of the facts by the individual, or with his consent; 5. The truth of the matter published does not afford a defense; 6. The absence of " malice" in the publisher does not afford a defense. The two remedies are: 1.An action of tort for damages in all cases; 2.An injunction, in perhaps a very limited class of cases. Obviously, the right to privacy is closely connected to publication. It is descript as a law term. They importance of individual and community is highlighted. The right to privacy is a kind of right they have.

However, is the right to privacy only meaningful to individual? The answer is no. With increasing knowledge of privacy, some scholars have been awareness of the importance of society in learning and defining the right to privacy. Solove [2006, p.483] has stated that privacy cannot be understood independently from the society in his law review. Similar thoughts can also be found in other resources. For example, Cohen [2000] summarizes three debates on the right to privacy: 1. Ownership of a certain kind of information; 2.Freedom of choice and its necessary preconditions; and 3.The

substantive value of personally identified information. He states that the costs of privacy are borne by society, or other individuals.

Privacy is more than a pure legal concept [Onn et al., 2005]. It has various meanings in psychological, social and political fields, and privacy causes different harms in different situations. Different concepts related to the right to privacy are summarized from different views, and one of these descriptions can be taken as an example [Onn et al., 2005, p.12]:

The right to privacy is our right to keep a domain around us, which includes all those things that are part of us, such as our body, home, thoughts, feelings, secrets and identity. The right to privacy enables us to choose which parts in this domain can be accessed by others, and control the extent, manner and timing of the use of those parts we choose to disclose.

The definition of the right to privacy is a generic level description. According to this definition, privacy is more like a right for people to choose, and it is tightly connected to people's permission and willing.

Another great contribution is given by professor Solove [2002] is his theory on conceptualizing privacy concepts and privacy issues. Solove proposed that privacy, or the right to privacy, is not a unitary concept. It has diversity across different situations. Instead of finding "family resemblances" from all privacy concepts, it would be easier to draw from a common pool of similar elements. Solove divided privacy rights into 4 types: information collection, information processing, information dissemination and invasion. A concept pool is provided as a strong support to analyst. The mind of concept pool has heavily influenced people's explosion on privacy protection. The definition of privacy and related concepts was debated over and over, and becoming increasingly clear. People's pursuit of privacy is still going on.

Although the concept of privacy never reaches a consensus conclusion, most researchers agree on its importance. The right to privacy kept drawing people's attention after it appeared. James Rachels stated in his paper "*Why privacy is important*" that privacy is a precondition for controlling people's various relationships that they value [Rachels, 1975]. It implies the features of privacy: valuable and powerful. However, these descriptions are obviously not clear and accurate enough. To further understand privacy, it is necessary to figure out in details what are "those things that are part of us".

2.2. Data privacy

Data privacy has no global, consistent definition. Even though, it frequently appears in literature and legal documents, such as "European Data Protection Law" [European

Data Protection Law, 2014] and "New Privacy Legislation" [McCormick and Michelle, 2011]. The first edition of European Data Protection Law was published in 1998. Now, it has been gradually accepted by European Union member states and becomes influential globally. In European Data Protection Law, the description of the right to data protection and personal data is respectively close to data privacy and personal information (PI), which will be discussed in this thesis. European Data Protection Law states that the data protection is "a right to protection against the collection and use of personal data", and it "forms a part of the right to respect for private and family life, home and correspondence" [McCormick and Michelle, 2011, p.14].

In another resource, "Privacy Engineer's Manifesto" [Dennedy et al., 2014, p25], data privacy is "a derivative of the substantive right to privacy in that it is about data that has been created about an individual a. by him- or herself, b. by others through observations and analysis, or c. by the consumption or processing (i.e., use) of that data about an individual by others." Compared to other definitions of privacy, the key point in this definition is data. Data privacy could be simply defined as a kind of special privacy, whose medium is data. Additionally, data is especially related to personal identifiers. In this thesis, the data privacy mainly means the privacy closely related to data in social media sites and applications, stored in service providers' servers, and distinguished from other privacy forms, such as the one in a bank or a police system.

The right to privacy changes with technological progress. With network and digital devices getting gradually integrated into people's daily life, an increasing number of data privacy disclosure events happen to users, organizations, and enterprises. Plenty of social network websites and web applications are accepted by users and change people's communication means, such as Facebook, Instagram[2], Twitter[3], LinkedIn[4], QQ, Wechat, Weibo[5] and so on. Social network enter people's life, so does data privacy. However, data privacy gets less attention than it deserves. When people enjoy the convenience and happiness brought by social network, they hardly notice the privacy policies on the sign up page and the risks behind the products and services. Most users finds privacy boring until their own self-interests are compromised [Dennedy et al., 2014]. Due to the highly developed information science and technology, digital privacy disclosure events are not far from people's life. Data privacy is not only crucial but also worthy deserving more attention from every social network services' practitioner and user.

[2] Instagram is an online mobile photo-sharing, video-sharing and social networking service.

[3] www.twitter.com

[4] www.linkedin.com

[5] Three popular social media applications in China.

2.2.1. Personal identifiable information

What shall data be like? There are plenty of data transferring on the Internet every moment, are they all under the protection of privacy? The answer is no. When discussing data privacy, some terms, such as personal data, personal information (PI) and personal identifiable information (PII), are often mentioned. They are all the data that people care about and shall be protected. This thesis will not distinguish the difference among these terms. PII will be used in this thesis for discussion. Traditionally and literally, PII is information that directly identifies an individual, or any other anonymous information which could be surely related to only one single person when all of those are combined. For example, one social security number can only identify one person, and it is a kind of PII. When information in age, blood type and birthday is combined, it is possible to identify one single person. They are PII, too. In addition, a person's name, age, gender, phone number, national even position, nickname, etc. is PII. The information is "those things that are part of us". In European Data Protection Law, there is a definition of personal data, which has similar meaning to personal information [European data protection law, 2014, p.36]:

Data are personal data if they relate to an identified or at least identifiable person, the data subject.

A person is identifiable if additional information can be obtained without unreasonable effort, allowing the identification of the data subject.

Authentication means proving that a certain person possesses a certain identity and/or is authorized to carry out certain activities.

There are special categories of data, so-called sensitive data, listed in Convention 108 and in the Data Protection Directive, which require enhanced protection and, therefore, are subject to a special legal regime.

Data are anonymized if they no longer contain any identifiers; they are pseudonymous if the identifiers are encrypted.

In contrast to anonymized data, pseudonymous data are personal data.

This is a more rigorous definition of personal identifiable information from a legal standpoint. The International Organization for Standardization (ISO) is an international organization and promotes worldwide proprietary, industrial and commercial standards. Similar definition can be found in ISO standards. The ISO/IEC 29100:2011 and ISO/IEC 29101:2013 [ISO 29101, 2013] definite a privacy framework and a privacy architecture framework respectively, which are both related closely to personal identifiable information. ISO/IEC 29100:2011 provides a privacy framework which [ISO 29100, 2011, p.1]:

- *specifies a common privacy terminology;*

- *defines the actors and their roles in processing personally identifiable information (PII);*
- *describes privacy safeguarding considerations;*
- *provides references to known privacy principles for information technology.*

ISO/IEC 29100:2011 is applicable to natural persons and organizations involved in specifying, procuring, architecting, designing, developing, testing, maintaining, administering, and operating information and communication technology systems or services where privacy controls are required for the processing of PII.

In ISO 29100:2011, more than 20 items are listed as examples, showing what personal information or personal identifiable information can be like [ISO 29100, 2011, p.8]. There are not only items which are commonly known as personal information, such as name, date of birth, gender, national identifiers, or ethnic origin, etc.

Examples
Age or special needs of vulnerable natural persons
Allegations of criminal conduct
Any information collected during health services
Bank account or credit card number
Biometric identifier
Credit card statements
Criminal convictions or committed offences
Criminal investigation reports
Customer number
Date of birth
Diagnostic health information
Disabilities
Doctor bills
Employees' salaries and human resources files
Financial profile
Gender
GPS position
GPS trajectories
Home address
IP address
Location derived from telecommunications systems
Medical history
Name
National identifiers (e.g., passport number)
Personal e-mail address
Personal identification numbers (PIN) or passwords
Personal interests derived from tracking use of internet web sites
Personal or behavioural profile
Personal telephone number
Photograph or video identifiable to a natural person
Product and service preferences
Racial or ethnic origin
Religious or philosophical beliefs
Sexual orientation
Trade-union membership
Utility bills

Figure 2-1: Example of attributes that can be used to identify natural persons [ISO 29100, 2011]

PII is a broad concept. For information owners, the importance of different kinds of PII shall be different. Telling if a piece of personal identifiable information is harmful or not would be another discussion. A word in European Data Protection Law informs the difference, “sensitive”. Some forms of PII are additionally considered “sensitive,”, because these PII can easily cause harm or discriminate against someone. Different cultures have different standards on sensitive PIIs, some common examples are shown as following [Dennedy et al., 2014, p.31]:

- *Information about an individual’s medical or health conditions*
- *Financial information*
- *Racial or ethnic origin*
- *Political opinions*
- *Religious or philosophical beliefs*
- *Trade union membership*
- *Sexual orientation*
- *Information related to offenses or criminal convictions*

Obviously, the examples above are more likely harmful if they are disclosed. They are the information which people try to protect from others. Everyone takes care of these information no matter they belong to themselves or others. However, for a certain system, it is hard and unnecessary to distinguish if the data is sensitive or not. PII, as a term, “*applies to those commercial entities that collect data that can be reasonably linked to a specific consumer, computer, or other device*” [Federal Trade Commission, 2014]. PII shall be the minimum unit in this thesis.

2.2.2. Privacy threat

Literately, privacy threat is something or actions which threats the right to privacy. There is not offical definition for privacy threat. However, privacy threat shows properties. Andreas Pfitzmann and Marit Hansen [2010] stated 6 privacy properties. They are: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. Deng, et al.[2010] refined these privacy properties and extended them to 8 privacy properties: unlinkability, anonymity, pseudonymity, plausible deniability, undetectability and unobservability, confidentiality, content awareness, and policy and consent compliance [Deng, et al., 2010]. From these 8 privacy properties, 7 privacy threat types are defined. They are: linkability, identifiability, non-repudiation, detectability, disclosure of information, content awareness, and policy and consent non-compliance. All privacy threats in the LINDDUN methodology are supposed to belong to one or plural privacy threat types above. Privacy threat is the thing that all privacy related practitioners want to keep away from their PII.

In order to explain all privacy properties, **item of Interests** (IOI) shall be introduced firstly. IOI means information on an individual, which could cause a privacy issue, such as subject, message, action, and so on. Some IOIs can be personal identifiable information.

The **unlinkability** [Pfitzmann and Hansen, 2010] of two IOIs means, an attacker cannot make sure if these two IOIs are linked or not. For example, to a stranger, a random user identifier and a random password have unlinkability, but two messages from the same user are linkable.

The **anonymity** [Pfitzmann and Hansen, 2010] of a subject means that in a group of subjects, an attacker cannot make sure which one is this subject. For example, in an online chat room, every user shows the same id when they chat, then every user in this chat room has anonymity. **Pseudonymity** means using pseudonym. A pseudonym is an identifier which a subject uses instead of real name. For instance, nick name in some social networks.

The **Plausible Deniability** [Roe et al., 1997] is the opposite effect of non-repudiation. It means there is no evidence to prove the concurrence of an event or action. For example, if a system does not record logs or messages, it cannot be determined that if this system was running at a passive moment.

The **undetectability** [Pfitzmann and Hansen, 2010] of an IOI means an attacker cannot make sure if this IOI exists or not. For example, in a social network website, an attacker cannot make sure if one certain phone number exists in database or not. The **unobservability** of the IOI has two points. Firstly, unobservability means the IOI is undetectable by attackers. At the same time, every subject involved in this IOI is anonymous to other subjects. For example, in a private chat room, all messages and user information cannot be detectable by other users. If a nickname is not unique and not verified, the information of every user in this chat room has unobservability.

The **confidentiality** means protecting authorized restrictions about information access and disclosure [McCallister et al., 2009]. For example, people transfer encrypted messages, or add access control to a database which contains private information. Confidentiality is known as a security property. It refers closely to preserve privacy properties, so confidentiality is also a privacy property.

The **awareness** has various concepts. Generally, it is explained as people’s a kind of perception, comprehension, projection or understanding to a certain objective. This

objective can be a set of elements in an environment among a period of time[Endsley, 1995]. It also can be status of a system [Sohlenkamp, 1998] or activities of other people, even themselves [Dourish and Bellotti, 1992].

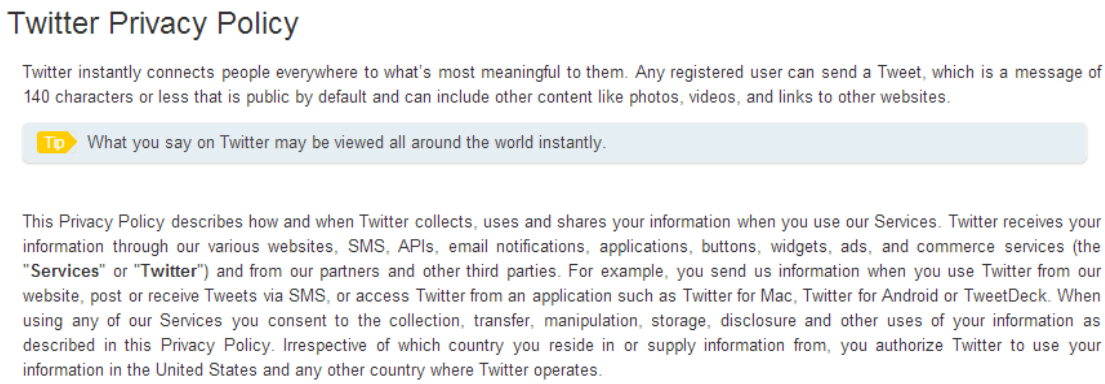
The **compliance** means a kind of agreement that the data subject allows others to process their personal data. All data shall be processed under the users' willing and permissions [European, 1995]. For example, in Facebook's privacy policy, there are items which explain what kind of data Facebook collects from users, and what kind of data shall be public. Users need to know enough details and agree with this privacy policy before they use Facebook service.

2.3. Privacy Protection In Software Development

Privacy is meaningful not only to people, but also to software development. All developers and engineers try to improve their system and make it trustworthy. In software development lifecycle, the earlier problems are found, the less they cost [Wiegers, 2003]. Privacy threat is supposed to be found and solved as early as possible to protect PII. The question is, when and how to deal with privacy problems in a software development lifecycle.

2.3.1. Privacy policy

A privacy policy is "a statement or a legal document (privacy law) that discloses some or all of the ways a party gathers, uses, discloses and manages a customer or client's data". [McCormick and Michelle, 2011] In ISO standards: "Privacy policy overall intention and direction, rules and commitment, as formally expressed by the PII controller related to the processing of PII in a particular setting." [IS0 29101, 2011] Privacy policy is a means of protection wildly used in software development. It is an official statement and a promise of privacy from service providers. Figure 2-2 is a screenshot of privacy policy presented on Twitter's privacy policy page. Twitter's privacy policy explains how and when Twitter collects, uses and shares users' information. It lists Twitter's methods to collect information in detail, and announces that users who get Twitter's service have authorized Twitter to use their information. Similar descriptions about privacy policy could be found on popular social network sites, like Facebook, LinkedIn, etc., and these descriptions partly present functions and features of a privacy policy.


Twitter Privacy Policy

Twitter instantly connects people everywhere to what's most meaningful to them. Any registered user can send a Tweet, which is a message of 140 characters or less that is public by default and can include other content like photos, videos, and links to other websites.

Tip What you say on Twitter may be viewed all around the world instantly.

This Privacy Policy describes how and when Twitter collects, uses and shares your information when you use our Services. Twitter receives your information through our various websites, SMS, APIs, email notifications, applications, buttons, widgets, ads, and commerce services (the "**Services**" or "**Twitter**") and from our partners and other third parties. For example, you send us information when you use Twitter from our website, post or receive Tweets via SMS, or access Twitter from an application such as Twitter for Mac, Twitter for Android or TweetDeck. When using any of our Services you consent to the collection, transfer, manipulation, storage, disclosure and other uses of your information as described in this Privacy Policy. Irrespective of which country you reside in or supply information from, you authorize Twitter to use your information in the United States and any other country where Twitter operates.

Figure 2-2 Privacy policy of twitter [6]

Privacy policy has a wide variety. Most countries have their own legislations and guidelines about who is covered, what information can be collected, and what it can be used for. Except for the data usage statement, the exact contents of a privacy policy will depend upon the applicable law and may need to address requirements across geographical boundaries and legal jurisdictions. The content of privacy policy can be various depending on system type. In general, a privacy policy consists of following items [Dennedy et al., 2014, p.80]:

- *Local and international legal, jurisdictional, and regulatory necessities, depending on the scope of the enterprise*
- *Organization or business requirements*
- *Permission for the marketing–customer relationship for management or business intelligence*
- *Brand identity*
- *Industry standards*
- *Usability, access, and availability for end users of information systems*
- *Economic pressure to create value through efficient sharing or relationship building*
- *Enforceability and compliance*
- *Ethical obligations*
- *Realistic technology capabilities and limitations*

Items above are common factors of a privacy policy. A complete privacy policy can contains plenty of content, even some of them are not related to privacy. A privacy policy not only presents the expectations from service providers to users, but also reflect strategies service provider use to solve privacy issues. It shows service provider's attitude to privacy in many perspectives.

As an official statement of a business enterprise or of a web service provider, privacy policy has the responsibility to help the enterprises to avoid potential legal issues. On

[6] This page is available as: https://twitter.com/privacy

the other hand, it shall be helpful to protect the users from potential personal information disclosure. Privacy policy plays an important role between social network and web application service providers and their users. Every mature enterprise shall have sufficiently consideration about the content of their privacy policy.

2.3.2. Microsoft privacy guideline

The Microsoft privacy guideline [2008] provides basic privacy concepts, privacy scenarios and rules to help software engineers and developers to build a privacy secure system. It offers privacy guidelines on a generic level. It would be useful for software engineers and developers who first time consider about privacy issues for their system. In first half part, Microsoft privacy guideline presents definitions about privacy and data types. It shows what privacy is, what privacy does, and what kind of data shall be protected. Then, in the other part, it provides nine privacy scenarios and some relevant rules to suggest users. It includes [2008]:

1. *Transferring PII to and from the user's system;*
2. *Storing PII on the user's system;*
3. *Transferring anonymous/pseudonymous data from user systems*
4. *Installing software on a user's system;*
5. *Deploying a web site;*
6. *Storing and processing user data at the company;*
7. *Transferring user data outside the company;*
8. *Interacting with children;*
9. *Server deployment.*

Scenario 2: *Storing PII on the user's system* is taken as an example. Firstly, Microsoft privacy guideline lists three possible examples in scenario 2: 1.Storing the user's contacts; 2. Caching Web pages that contain PII; 3.Storing PII in cookie. Secondly, all guides and suggestions are started with either "must" or "should". For example, "Users must be able to review and edit stored PII they entered", or "Users should be able to control whether PII is stored, and delete any PII stored on the user's system, including hidden PII". There are totally 8 guides in scenario 2.

The Microsoft privacy guideline is a useful tool for developers, engineers and analysts. The suggestions in the Microsoft privacy guideline contains plenty of situations which developers might meet in practical project. All the rules and suggestions are like privacy requirements, and they are easy to understand. However, it has limitations. All rules are made based on a fixed way, which cannot apply for some specific systems. Nine scenarios can apply for most situations, but not all. Besides, a combination of guidelines and rules is not a structured approach. The Microsoft privacy guideline is a good standard to verify the correctness of other methods. All privacy requirements, which can

be related to the rules in the Microsoft privacy guideline, are regarded as correct privacy requirements.

2.3.3. ProPAn

ProPAn [Beckers, et al., 2014] is a structured approach for analysts to semi-automatically identify privacy threats. Its workflow is shown in Figure2-3. The ProPAn consists of four steps: Draw context diagram and problem diagrams, add privacy requirements to model, generate privacy threat graphs and analyze privacy threat graphs. The ProPAn methodology is an approach to identify privacy threats from functional requirements and privacy requirements. The frist two steps are regarded as preparation steps. With the input of the first two steps, threat graphs will be automatically generated in the third step, and privacy threats can be analyzing in last step.

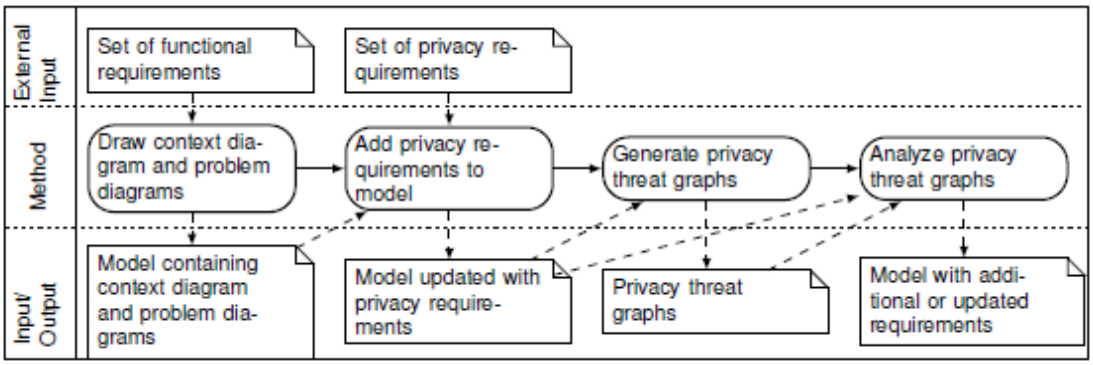


Figure 2-3 Illustration of the ProPAn method [Beckers, et al., 2014]

The ProPAn methodology is strictly not for privacy requirements analysis. It is used to identify privacy threats. It means that, before the ProPAn method is applied, problems are clear, and requirements are ready. It is good for experts and analysts to find hidden privacy threats in a system. However, for a normal user with limited privacy knowledge, it is hard to elicit privacy requirements directly form a system. The ProPAn has its limitation on its usage.

2.3.4. Summary

In this chapter, three privacy protection means are introduced and discussed. Privacy policy is a widely used means to protect privacy. For every application which collects users' information, there shall be a privacy policy to inform users and avoid potential legal issues for the enterprise. The Microsoft privacy guideline contains plenty of useful suggestions and rules about privacy protection in software development. It is a combination of people's experience on protecting privacy. But it is not a constructive approach. The ProPAn is a threat identifying method, which has detailed procedures. It

is a good example of threat identifitition methods. However, it does little help on privacy requirements eliciation. All three privacy protection means have advantages and disadvantages.

2.4. Introduction to LINDDUN

The LINDDUN methodology is designed to address the privacy threats in a system for software engineers. It is helpful and practical to identify privacy threats and elicit privacy requirements. It can be applied at different stages of a project, such as at the architecture stage, at the requirement stage, or with a system sketch. LINDDUN consist of a summary of privacy concepts, several sets of check lists and privacy-enhancing technologies list.

The LINDDUN methodology [Deng et al., 2010] is inspired by a security framework, STRIDE [2002], and evaluated and supplymented by Kim Wuyts in 2015. Kim Wuyts and her group made two empirical expirements on LINDDUN. The improvements are mainly based on the experiments results. In the same year, Kim Wuyts and Wouter Joosen publicshed a tutorial for LINDDUN Framework 2.0 [Wuyts and Joosen, 2015], which greately improved the usability of LINDDUN.

The core contributions of LINDDUN are the conceptional framework and the technical process method. Firstly, LINDDUN integrates the definitions of privacy threats and privacy properties, and presents the related check lists, which are helpful to check potential threats. The name, LINDDUN, is an abbreviation of seven privacy threat types: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, content Unawareness, as well as policy and consent Noncompliance. On the other hand, LINDDUN presents a guide line to scan a system privacy claws. The LINDDUN methodology offers a new approach for privacy requirement analysts and engineers to model privacy threats and elicit privacy requirements from a system.

3. A privacy threats model – LINDDUN

3.1. Concepts and definitions in LINDDUN

Danezis[2008] puts forward that privacy can be devided into hard privacy and soft privacy. **Hard privacy** has the similar goal with data minimization, assuming that the third party is untrustable, such as the third party service provider or advertiser. A system of hard privacy tries to reduce the possible to ”trust” other entities. **Soft privacy**, on the contrary, means that users provides data to the thrid party as much as they need, and trust them. It is under the assumption that the data controller is resposible for data protection. Besides, **privacy mitigation strategy** means strateties or measures which can mitigate privacy threats. **Mitigation technology** means technical means to mitigate privacy threats.

3.2. LINDDUN procedures

LINDDUN is a privacy threat analysis methodology that supports analysts to elicit privacy requirements from a system stretch. There are 6 primary procedures in LINDDUN to elicit privacy requirements. Figure3.1 shows a step-by-step overview of the LINDDUN methodology using a simple social network system as a running example. The approach divides privacy requirements analysis into two phases. The first phase is the identification of privacy threats, and it is conducted in three steps, which are considered as the core steps of LINDDUN. The rest three steps are solution-oriented and aim at translating the threats, which have been identified, into viable strategies and solutions that can mitigate the threats.

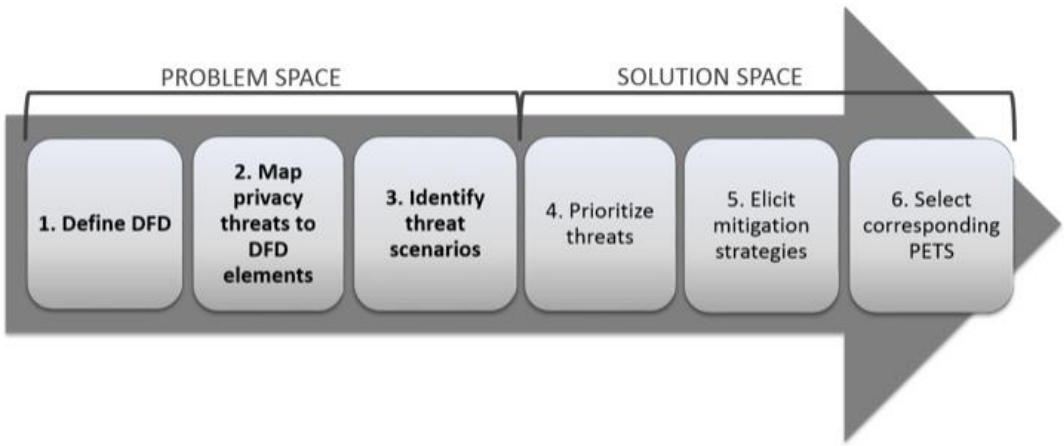


Figure 3.1 The LINDDUN methodology steps [Wuyts, 2015]

Before using LINDDUN methodology, some basic knowledge in LINDDUN, which are introduced in previous chapters, such as privacy threat, privacy property and item of interests are necessary. At the same time, the analyst needs to be familiar with the

system. When all preparations are ready, the analyst can start to analyze the system step by step. The first step is to model a data flow diagram (DFD) of the system. Based on the analyst's comprehension to the system, a data flow diagram is created as output of step 1. In the second step, the analyst needs to locate all privacy threats associated with every data flow diagram element with the help of privacy threat type. The goal in this step is to avoid of missing any privacy threats, and remove reduplicative privacy threats. A table with key privacy threats marks is created as step 2's output. The third step is to elicit and document threats. LINDDUN has a strong support in this step, called threat tree. All threat tree are up to date by LINDDUN researchers. The analyst uses the threat tree to get specific privacy threats from step 2's table. The threat tree is used as a check list. Once the analyst gets one specific privacy threat, he documents this privacy threat as a misuse case. After this step, a list of misuse cases is supposed to be output of this step. Problem-oriented steps stop here. The forth step is to prioritize all privacy threats. After all problem-oriented steps, there shall be plenty of privacy threats which need to be considered according to the system's size. Prioritization is necessary to figure out which privacy threats are more critical and shall be focused. In next step, a mitigation strategy tree is given by LINDDUN methodology to help the analyst to find strategies for every privacy threat found in previous steps. Finally, in the last step, a privacy enhancing technology list is given to support all strategies. According to these strategies and technologies, a list of privacy requirements can be produced in detail.

3.2.1. Model Data flow diagram (DFD)

The first step of the LINDDUN methodology is to model a data flow diagram for the system. The data flow diagram (DFD) is a visual tool to describe logic models and expresses data transformation in a system [Li and Chen, 2009]. The data flow diagram can illustrate details of the functions a system possesses. Data flow and process can be shown simply and clearly in a DFD. Data flow diagram, compared to other diagrams, like use case diagram or sequence diagram, has its advantages. In privacy requirements elicitation, data shall be the main concern. DFD has advantages especially on documenting data flows or exploring a new high-level design in terms of data flow [Craig Larman, 2005]. DFD can express all stored data and transferring data in a system. It is an appropriate tool for privacy analysis.

A DFD consists of combined by 4 main elements: data input, data flow, data process and database. Figure 3.2 is a sample DFD of social network application. There are data transmission between users and portal process, portal process and service process, and service process and a social network database. There is only one database and one data input in this case.

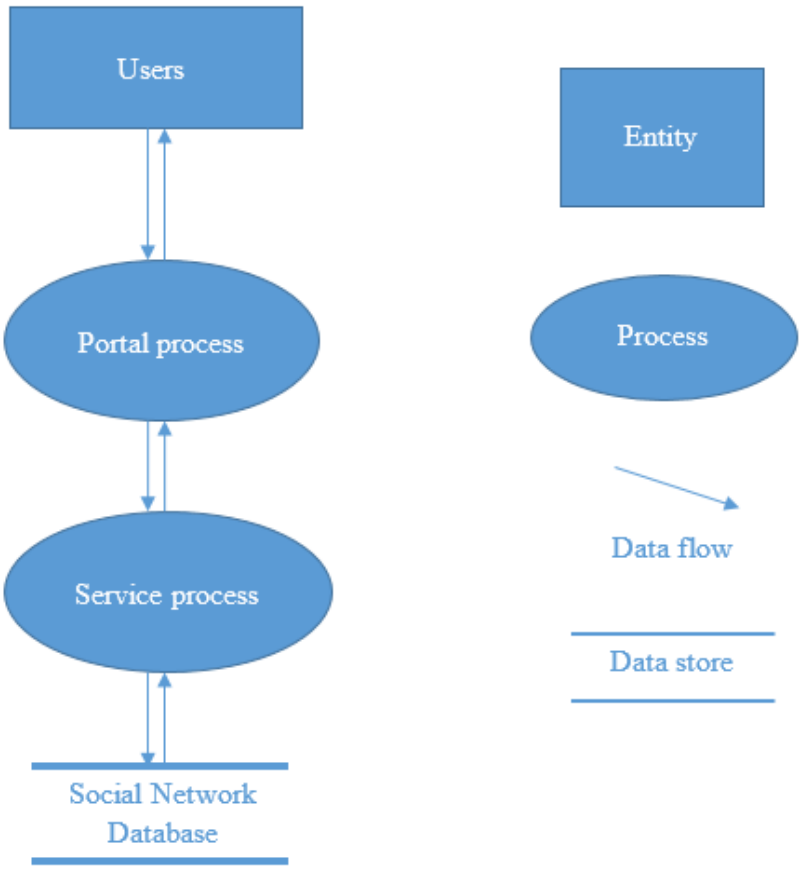


Figure 3.2 A sample of data flow diagram [Deng et al., 2010]

The DFD has its own rules on how to create a standard DFD. The granularity of a DFD has a flexible range [Li and Chen, 2009]. A process, also called activity, is the transformation of data. It can be further decomposed to form more detailed sub-process. For instance, a DFD of a social network can be simply like the example in Figure 3.2. It can also be made in details if every service process is regarded as one separated process, such as upload process, share process, add friends process, and so on. For another instance, a system with two entities shall be more complex than the system with only one entity. More elements a DFD has, more complex the DFD is. An analyst shall make a clear and correct DFD for a good start. How to create a good DFD? How much details shall a good DFD include? About DFD, there is little explanation in Deng and Wuyts' thesis. This shall be a valuable point in the following case study.

3.2.2. Map privacy threats to DFD elements

The second step is to map privacy threats to DFD elements. In LINDDUN framework, all privacy threats are catalogued into 7 types. They are respectively Linkability, Identifiability, Non-repudiation, Detectability, Information Disclosure, Content Unawareness and Policy/consent Noncompliance. Every DFD element is regarded as an independent unit. LINDDUN has done some effort to reduce user's workload. The

crosses in Table 3.1 mark potential privacy threat points that needs to be cared about, and the blanks in table means the part that shall not be considered. For example, linkability of entity is one point which is supposed to be considered as a privacy threat. If two accounts have similar profiles, they might be from the same entity. For another example, all DFD elements are parts of the system except for entity, so only entity has content awareness in the system. Content unawareness of data flow, data store or process does not need to be analyzed in LINDDUN methodology. Further discussion about these marks will be talked in next step.

THREAT CATEGORIES	Entity	Data Flow	Data Store	Process
Linkability	×	×	×	×
Identifiability	×	×	×	×
Non-repudiation		×	×	×
Detectability		×	×	×
Information Disclosure		×	×	×
Content Unawareness	×			
Policy/consent noncompliance		×	×	×

Table 3.1 Map threats to DFD elements [Deng et al., 2010]

In a practical case, every DFD element could have multiple units. In Figure 3.2 sample, there are 1 entity, 6 data flows, 1 data store and 2 processes. For every entity, there are 3 potential privacy threat points, and for every other element, the number increases to 6. Then, there are 3+18+6+12 = 39 potential privacy threat points in total. Every point is one situation which should be analyzed in following steps, and might lead to several privacy threats. In order to improve the productivity, unnecessary and repetitive points shall be removed. Judgment on whether one case matches on one privacy threat shall critically obey the definitions of these privacy threats, which has been introduced in previous chapters. A short guide on judgment is given by LINDDUN as a reference:

- ***Linkability*** *(L) occurs when one can sufficiently distinguish whether 2 items of interest (IOI, such as requests from a user) are related*
- ***Identifiability*** *(I) occurs when it is possible to pinpoint the identity of a subject (e.g., a user)*
- ***Non-repudiation*** *(Nr) occurs when it is possible to gather evidence so that a party cannot deny having performed an action*
- ***Detectability*** *(D) occurs when one can sufficiently distinguish whether an IOI exists, e.g., in a system*
- ***Disclosure of information*** *(Di) is the exposure of information to individuals who are not supposed to have access to it*

- ***Unawareness*** *(U) occurs when the user is unaware of the information he is supplying to the system and the consequences of his/her act of sharing*
- ***Non-compliance*** *(Nc) occurs when the system is not compliant with the (data protection) legislation, its advertised policies and the existing user consents*

Besides above criteria, LINDDUN methodology puts forward other means to minimize the number of privacy threat points, called making assumptions. Making assumptions means when an analyst use LINDDUN methodology to analyze a system, he/she shall make some assumptions according to specific conditions of that system. These assumptions always influence privacy threats elicitation more or less. For example, linkability and identity of entity are only applicable when the social network system is anonymous. If the system is not anonymous, there is no need to protect users' identity information. "The system is anonymous" shall be one useful assumption, and the marks of linkability and identity of entity shall be removed. In LINDDUN official tutorial [Wuyts and Joosen, 2015], there is another important action: combine "X"s. This combine "X"s action is a part of making assumptions. The word "combine" reveals the essence of making assumptions: reduce the number of privacy threat points. Making assumptions determines which part of the system shall be ignored and which part deserves more attention. It greatly affects workload of all analysis processes and final result.

3.2.3. Elicit and document threats

This step contains two parts: eliciting threats and documenting threats.

Elicit threats: After mapping the privacy threats on the system, a table with marks shall be ready as the input of next step: eliciting threats. In this step, LINDDUN methodology provides a privacy threat tree catalog as a checklist to help analysts continue eliciting privacy threats from potential privacy threat points table. The privacy threat tree catalog is supposed to consist of all possible options for any pair of privacy threat and DFD element. For each potential privacy threat point in last step, there shall be at least one corresponding privacy threat tree for eliciting privacy threats.

The completeness of threat tree effects the completeness of LINDDUN methodology. Threat tree shall be up to date along with new privacy threats' discovering. Present version of threat tree has been improved by in Kim Wuyts [Wuyts et al., 2015]. Lastest version is available online[7].

[7] https://people.cs.kuleuven.be/~kim.wuyts/LINDDUN/

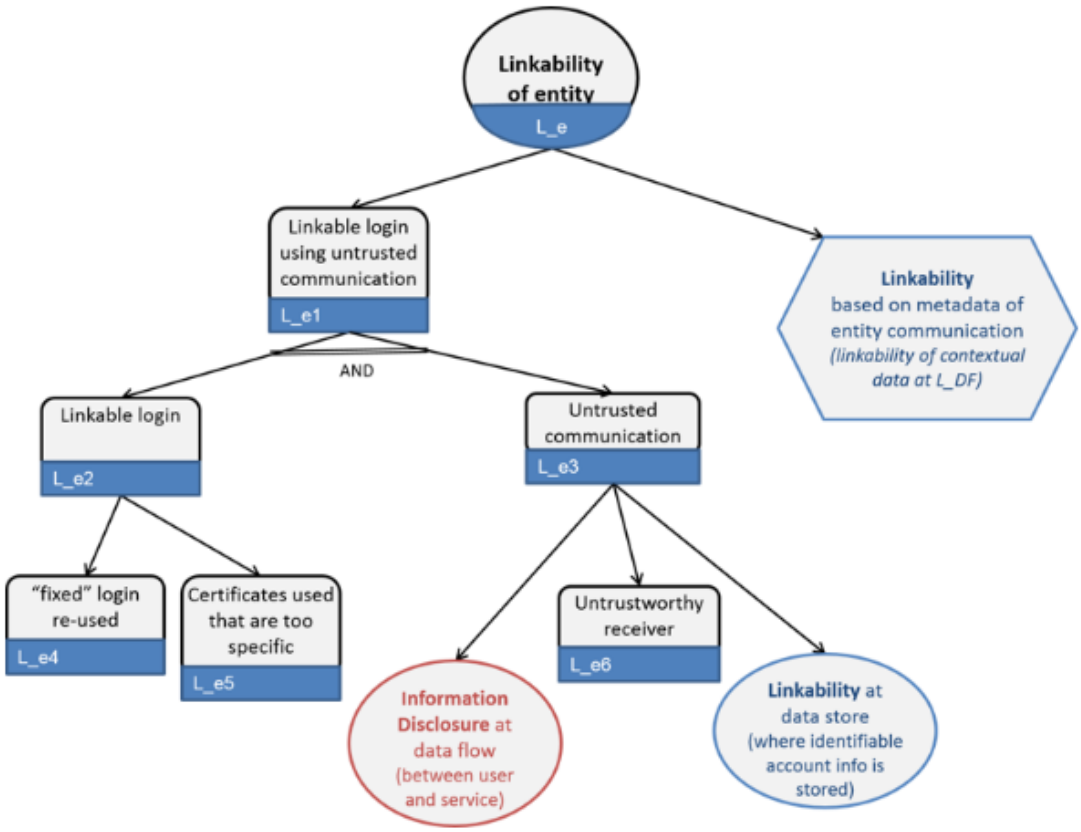


Figure 3.3 Linkability of an entity threat tree [Wuyts et al., 2015]

Figure 3.3 is a threat tree for linkability of an entity. The first round box is catalog title, called root node. Others are leaf nodes. There are two subtrees under this root node. The box on left side is a specific situation about how linkability of entity effects on a system. Any leaf node under this node is a further discussion about this situation. Besides problem descriptions, round box means this node refers to another threat tree. In this example, they are information disclosure at data flow and linkability at data store. They are root nodes in their own threat trees. Information Disclosure at data flow is a special node, whose edge is red in privacy threat tree catalog. It means that this node belongs to a security threat tree. Security threat trees are also listed in LINDDUN framework and might be used in some cases. The sexangle box on right side is a subtree, which also appears in another threat tree: Linkability of data flow. To avoid duplication, this subtree is only shown in linkability of data flow threat tree.

The threat tree can be used as a checklist in privacy threats elicitation. LINDDUN is designed to solve generic privacy problems, so all system shall be applicable to these treat trees. For an analyst, the only thing is excluding redundant items in threat tree. Analyst is firstly supposed to pick one threat tree, then, consider if the system has same or similar privacy threats with descriptions in leaf nodes. LINDDUN provides further

explanation under every privacy threat tree. Once a privacy threat is elicited, it needs to be documented in next step.

Document threats: Another part is to document threats. A threat template is necessary to document threats. The misuse case is recommended by LINDDUN to document threats. Misuse case can be used to express threats. Conversely the requirements can also be elicited from misuse cases. Misuse case is a suitable way to document threats once the threats are found. The proposed misuse case structure is shown in Table 3-2 with a brief explanation as below (optional fields are indicated with *) [Guttorm S. and Andreas L. 2001]:

"Summary" is a brief introduction to a threat. It describes a privacy threat with one or two sentences. "Assets, stakeholders and threats" means the people who might be threaten by this threat. "Primary misactor" means the one who cause this privacy threat, and "trigger" is the action made by misactor. "Basic flow" and "alternative flow" are the processes of privacy threat cause harm to system or users. "Preconditions" are assumptions made by previous steps. "Leaf node" and "root node" are used to located the options in privacy threat tree. These nodes are helpful to find mitigation strategies in further analysis. "DFD element" reveals the connections between privacy threat and data flow diagram. "Leaf node, root node and DFD element" shall be mandatory fields in this thesis.

Number and title

Summary	Provides a brief description of the threat.
*Assets, stakeholders and threats**	Describes the assets being threatened, their importance to the different stakeholders, and what the potential damage is if the misuse case succeeds.
Primary misactor	Describes the type of misactor performing the misuse case. Possible types are insiders, people with a certain technical skill, and so on. Also, some misuse cases could occur accidentally whereas other are most likely to be performed intentionally.
Basic flow	Discusses the normal flow of actions, resulting in a successful attack for the misactor.
*Alternative flow**	Describes the other ways the misuse can occur.
*Trigger**	Describes how and when the misuse case is initiated.
*Preconditions**	Precondition that the system must meet for the attack to be feasible.
*Leaf node(s)**	Refers to the leaf node(s) of the threat tree(s) the threat corresponds to.

*Root node(s)**	Refers to the root node(s) of the threat tree(s) that were examined for the threat.
*DFD element(s)**	Lists all DFD elements to which this threat is applicable.
*Remarks**	Although optional, related assumptions shall be mentioned here.

Table 3-2 A template for miause case [Guttorm S. and Andreas L. 2001]

After this step, all privacy threats shall be done and documented. That means all problem-oriented steps in LINDDUN methodology are finished. Next step is to solve these problems and get final target: privacy requirements list.

3.2.4. Threats prioritization

From this step, the goal is to solve privacy threats, rather than to find and define them. In order to solve these privacy threats, it is necessary to figure out how serious a threat is. A threats prioritization shall be made according to some measures. It is like a risk management in requirement engineering, but the privacy threat is concerned in this process. Methods on making prioritization is not mentioned in LINDDUN. The Open Web Application Software Project(OWASP) [2015], provides a methodology named OWASP Risk Rating Methodology (ORRM). It is taken as an option to do a prioritization.

The ORRM approach is based on standard methodologies and customized for application security. The most common risk model shall be:

Risk = Likelihood * Impact

Risk is a number that shows the risk of a threat. The higher the number is, the more important the threat shall be. Likelihood is a number from 0 to 9 that shows possibility of this threat. Impact is a number from 0 to 9 that means if this threat succeeds, how much damage it causes. The procedure of ORRM has 6 steps:

Step 1: Identifying a Risk

When a user starts to rate risks, these risks shall be identified and ready for the following work. In this case, all privacy threats shall be identified at the beginning. After step 3 of LINDDUN, all privacy threats have been elicited from the system and documented with misuse case template.

Step 2: Factors for Estimating Likelihood

In ORRM, the factor is an important concept to estimate severity. There are eight factors which influence the likelihood of one risk, including skill level, motive, opportunity, size, ease of discovery, ease of exploit, awareness and intrusion detection.

Not all factors are applicable in analysis of the risk of every threat. For different projects, selection of factors can be customized. Each factor has a set of options, and each option has a rating number from 0 to 9. For example, the first sentence is the meaning of this factor. Then, all options are followed with brief descriptions, and ordered from high to low. These option points are discontinuous integers. Only one option can be selected for one risk. In the end of step 2, the average points of all factors presents a likelihood level for each threat.

- *Skill level*

How technically skilled is this group of threat agents? Security penetration skills (9), network and programming skills (6), advanced computer user (5), some technical skills (3), no technical skills (1)

Step 3: Factors for Estimating Impact

Factors for estimating impact are similar to likelihood. There are eight factors which can influence the impact of one risk. They are loss of confidentiality, loss of integrity, loss of availability, loss of accountability, financial damage, reputation damage, non-compliance and privacy violation. Not all factors are applicable in analysis of the risk of every threat. For different projects, selection of factors can be customized. Each factor has a set of options, and each option has a rating number from 0 to 9. Reputation damage is taken as an example. There are descriptions to factors and options. These option points are discontinuous integers and ordered from high to low. Only one option can be selected for one risk. In the end of step 3, the average points of all factors presents an impact level for each threat.

- *Reputation damage*

Would an exploit result in reputation damage that would harm the business? Minimal damage (1), Loss of major accounts (4), loss of goodwill (5), brand damage (9)

Step 4: Determining Severity of the Risk

There are three levels to estimate likelihood and impact in ORRM as shown in Table 3-3. Point from 0 to 3, including 3, is rated LOW. Point from 3 to 6, including 6, is rated MEDIUM. Point from 6 to 9, including 9, is rated HIGH. All estimated risks shall be divided into these three levels, according to the average point of each risk.

Likelihood and Impact Levels	
0 to <3	LOW
3 to <6	MEDIUM
6 to 9	HIGH

Table 3-3 Likelihood and impact levels [OWASP, 2015]

Overall risk severity level is influenced by likelihood and impact levels. There are five levels for overall risk severity: critical, high, medium, low and note, as shown in Table 3-4. Finnaly, every risk, or privacy threat will get a severity level. Usually, risks above medium shall take more attention compared to other risks.

Overall Risk Severity				
Impact	HIGH	Medium	High	Critical
	MEDIUM	Low	Medium	High
	LOW	Note	Low	Medium
		LOW	MEDIUM	HIGH
	Likelihood			

Table 3-4 Overall risk severity levels [OWASP, 2015]

Step 5: Deciding What to Fix

There is none instruction in this step. The decision shall be made depends on the result and specific situations of the system.

Step 6: Customizing Your Risk Rating Model

This step makes this risk model more applicable to generic projects. However, there is no need to customize this model after prioritization is produced.

3.2.5. Elicit mitigation strategies

Eliciting mitigation strategy is a process in which privacy threats are turned into privacy threat mitigation strategies or privacy requirements. Privacy threat is not the final goal of the LINDDUN methodology, and specific solutions are more practical for software developers and other stackholders to handle with. In this step, LINDDUN provides a guide to lead analysts from the problem to the solution in theory step by step. It shall be highlighted that one privacy threat might be connected to a number of requirements, vice versa.

In LINDDUN, there is a taxonomy of privacy mitigation strategies, as shown in Figure 3-4. LINDDUN's authors divide privacy mitigation strategies into two types: concealing association and guarding association. Concealing association usually protects data by hiding it or faking it, and tries to make data ignored by attackers. According to the protected objective, concealing association has two deeper types: protect ID and protect data. Due to the variety of data content, protect data can be further classified. On the

other hand, guarding association is a more initiative method compared to concealing association. It also has two types as well. Guard exposure tries to protect data from exposure. Maximize accuracy makes people aware of data disclosure as early as possible. Every item in Figure 3-4 is related to a series of mitigation strategy technologies in Table 3-5. The taxonomy of privacy mitigation strategies is greatly helpful to find corresponding mitigation strategies for different privacy threats.

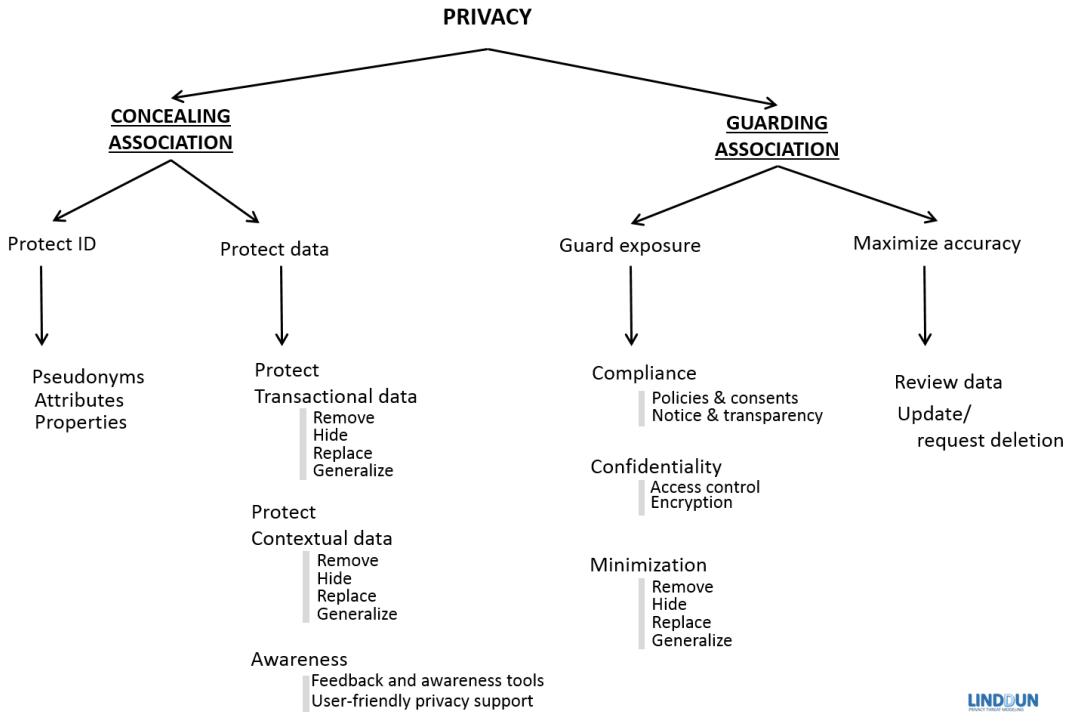


Figure 3-4 Taxonomy of privacy mitigation strategies [Wuyst, et al., 2015]

Mitigation Strategy	LINDDUN Threat Tree
Protect ID	L_e, I_e
Protect data	
Transactional data	L_DF1, I_DF1
Contextual data	L_DF2, I_DF2, D_DF, NR_DF
Awareness	U_1
Guard exposure	
Compliance	NC
confidentiality	ID_DS, NR_DS, *_P
Minimization	L_DS, I_DS, D_DS
Maximize accuracy	
Review data	U_2
Update/request deletion	NR_DS3

LINDDUN

Table 3-5 Mapping of mitigation strategies to threat tree [Wuyst, et al., 2015]

A table provided by LINDDUN shows how to find proper mitigation strategy for each privacy threat, as shown in Table 3-5. For instance, if a privacy threat refers to the unawareness of an entity, it shall belong to U_2 situation. The strategies could be maximize accuracy and review data. Every privacy threat from previous work shall be located in this table. It offer specific minds to solve privacy problems, and it leads to specific solutions, privacy enhancing technologies.

3.2.6. Select privacy enhancing technologies (PETs[8])

After 5 steps, a list of privacy threats and corresponding mitigation strategies shall be ready. Finding mitigation technologies is also a part of LINDDUN. This is also a part of LINDDUN methodology. There are common solutions or strategies to meet privacy requirements:

1. *Warn the user* could be a valid strategy for lower risk threats. However, user is the one who makes decisions. Only warning the user cannot sufficiently solve privacy problems.
2. *Removing or turning off the feature* can completely remove privacy threats. If the feature has more risks than benefits, the best action is to give up this feature. This also working for privacy problems. Nothing will be disclosed if user hasn't shared anything.
3. *Countering threats with either preventive or reactive privacy enhancing technology* is the most commonly used strategy to solve privacy issues.

In order to solve the practical problems, privacy enhancing technologies are necessary. The LINDDUN methodology provides a table which offers the relationship between mitigation strategies and privacy enhancing technologies. Privacy enhancing technology list is available and live update online. Considering that specific technology is not key point in this thesis, introduction to privacy enhancing technology is omitted. Analyst can find corresponding technologies in this table, according to strategy of each privacy threat. Then, translate these strategies and technologies into documented privacy requirements. So far, the procedure of applying LINDDUN to a system is finished.

[8] PET, Annual symposium on privacy enhancing technologies (PETs), available as: http://petsymposium.org/

4. Apply LINDDUN to the Rin-Tin-Tinder project: A case study

4.1. Purpose and hypothesis

The research objective of this thesis is to figure out how practically we can apply the LINDDUN methodology in privacy requirements analysis. In this chapter, the author will apply the LINDDUN methodology to the Rin-Tin-Tinder (RTT) system. The workshop result and the Microsoft privacy guideline will be involved as assists. The number of privacy threats and privacy requirements from the LINDDUN methodology and if most of them are consistent with other methods, such as the Microsoft privacy guideline, are two key points which shall be focused. The number of LINDDUN privacy threats will be compared with the number of workshop privacy threats. On the other hand, the Microsoft privacy guideline is a good reference to verify the correctness of LINDDUN privacy requirements. A privacy requirement is regarded as correct if there is a similar guide or rule in the Microsoft privacy guideline. Then, the objectives of the case study is twofold, and they are 1) the LINDDUN analysis provides a systematic way to identify privacy threats and to elicit privacy requirements for software applications, compared with the requirements elicitation techniques such as workshop; 2) the elicited requirements are consistent with the Microsoft privacy guideline.

4.2. Introduction to the case: Rin-Tin-Tinder

Rin-Tin-Tinder(RTT) is a social web application. It is a project from Demola[9]. The RTT team consists of 5 developers from university, one facilitor from Demola community, and one client from a Finnish company. I am involved in this project as a developer.

[9] Demola is an international organization that facilitates co-creation projects between university students and companies, either locally or internationally. Link to Demola: http://www.demola.net/

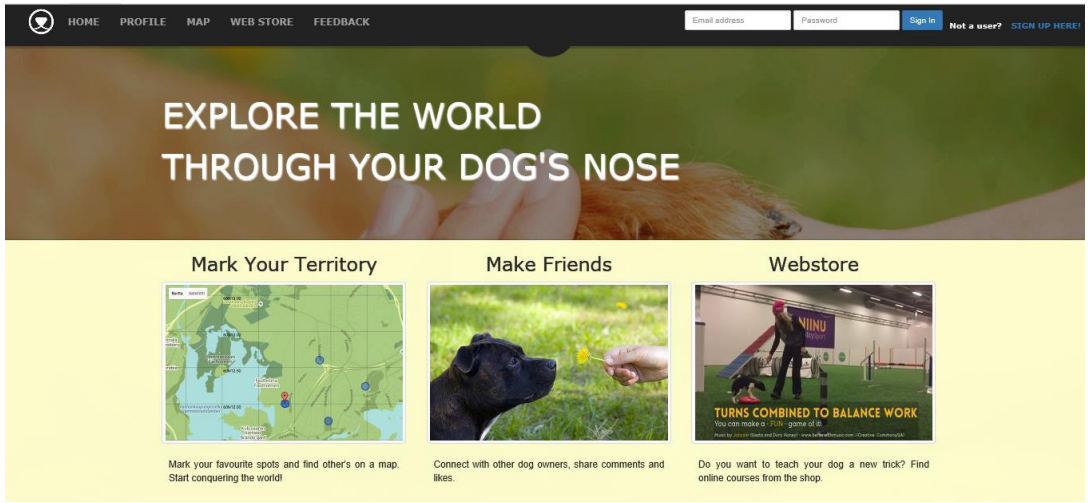


Figure 4-1 A screenshot of the Rin-Tin-Tinder web application

RTT is created for dog owners. The goal of RTT is to create a dog owner community. In this community, every dog owner can make friends with others, communicate with each other, get dog training instuctions from dog trainers or dog experts, and buy dog products online, including dog training courses. All these features are combined together to build a dog owners version Facebook. RTT is designed to be a social network website application. A demo of the RTT web application is available online[10]:

As a social network site, RTT has many features. A use case diagram presents most features in Figure 4-2. There are two kinds of users in the RTT system: Content provider user and normal user. Normal user, as main object of the RTT service, are supposed to be dog owners. They can use the sociality feature just like other social network websites, such as login, logout, edit profile, share status and Q&A. They can also import profile or pictures from other websites, buy dog online courses or other products in the RTT webstore, or check in on the map when they run their dogs. On the other hand, content provider user acts as an administrator of RTT system. They are from the RTT running team. Besides basic login and sociality services, content provider user shall push notifications and update web store products for normal user.

[10] Link to Rin-Tin-Tinder: http://rintintinder.herokuapp.com

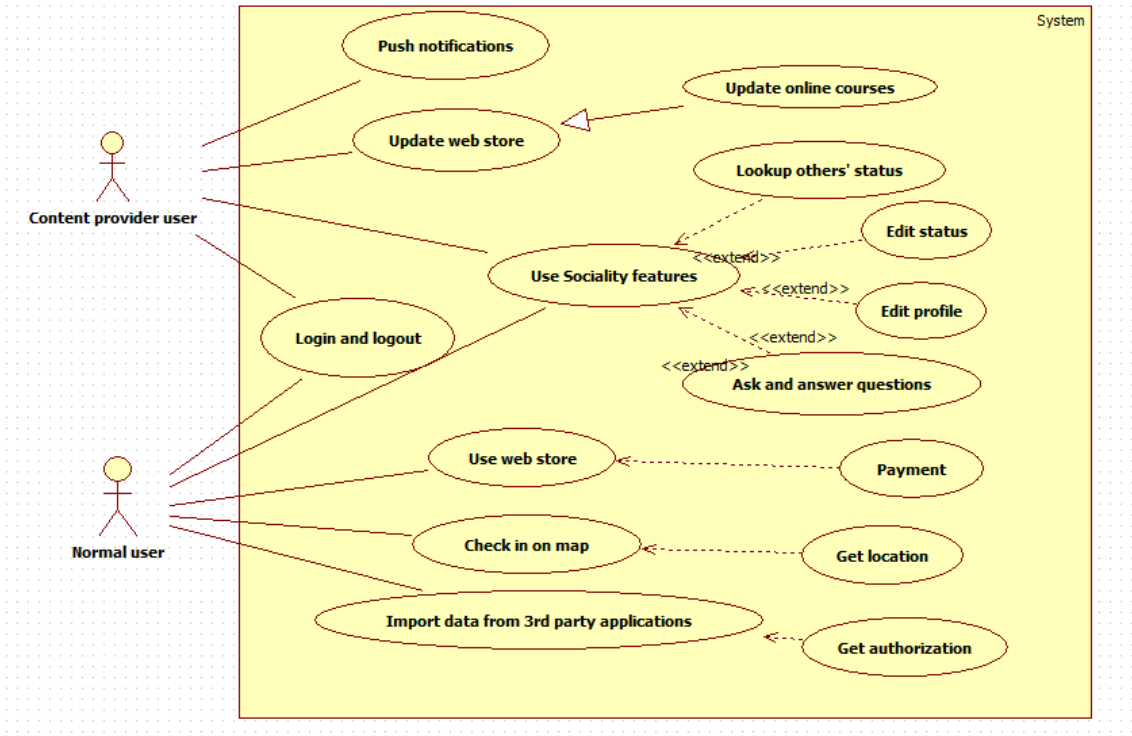


Figure 4-2 Use case Diagram of the Rin-Tin-Tinder web application

The developer team needs to consider about privacy issues before the system releases. While this thesis was writing, the RTT project was on its design stage. The structure of the RTT system is clear, but the coding work haven't completely done. It is a proper time to analyze privacy requirements for the RTT system. Besides, three developers of the RTT development group had workshop to identify privacy threats from the system, which is another reason that the RTT system is choosen.

4.3. Methods used without LINDDUN

4.3.1. Workshop

Before LINDDUN is applied for the RTT project, the development group did a workshop to identify privacy threats. A workshop is a meeting in which a group of people apply methodologies for a certain subject, in order to achieve a result. A workshop always needs a topic. It can be a speech given by several people, it also can be only a meeting for solving specific problems.

The author owned and facilitated this workshop. There are three participants from the RTT project development group in this workshop. One of them plays as scribe during the workshop. They are all undergraduates from University of Tampere, and their majors are all related to information science. Participants are requested to come up with as many privacy threats as possible for the RTT system. There are three participants involving in this workshop. They are developers from the RTT project group. They are

familiar with the RTT system, but they do not know LINDDUN. When the workshop is held, the RTT project is on its architecture phrase, which is a perfect timing to start analyzing privacy requirements. In this workshop, participants followed the author's suggestion, using data flow diagram as an assist.

The procedure of the workshop is as below:

1. Before the workshop starts, three participants are asked to create a data flow diagram of the RTT system, according to the RTT system features;
2. After the data flow diagram is done, brief introduction to privacy threat is provided to participants as basic knowledge for follwing steps;
3. Participants are required to identify privacy threats as many as possible in a two hours meeting;
4. Participants are required to sort privacy threats and abandon incorrect privacy threats.

4.3.2. Workshop result

After two hours workshop meeting, two output are produced. The first one is a data flow diagram, as shown in Figure 4-3. This data flow diagram has 2 entities, 7 processes, 13 data flows and 4 data stores. This is a simplified data flow diagram, using non-arrow lines. All data flows are not clarified clearly in this diagram.

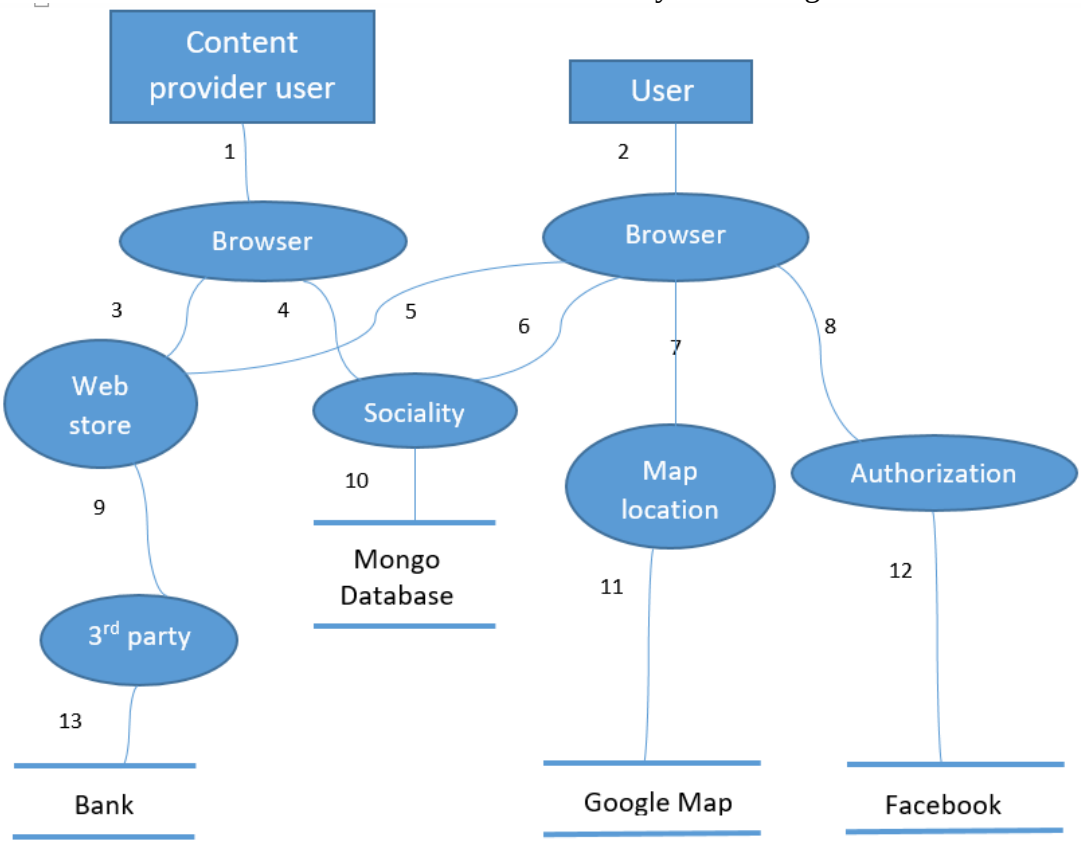


Figure 4-3 A data flow diagram by the RTT developer group workshop

Except for the data flow diagram, another result is a privacy threat list shown in Table 4-1. These threats are all put forward by participants with the help of DFD in Figure 4-3. It cost three participants around 30 minutes to create the DFD, and around 90 minutes to find 11 privacy threats for the RTT system. These threats are problems which might happen to the RTT system so far for the RTT developer group.

These privacy threats from workshop are not described with much detail. After the DFD is done, participants try to come up with all possible threats in every process, and record them all. All participants do not use any techniques except their brains. Finally, Table 4-1 is produced as another result of the workshop.

Threats	Misactor
W01: Infected browser is used.	**Skilled outsider**
W02: Hijack the RTT website, to make an infected version.	**Skilled outsider**
W03: Someone can get user's location, and use user's location data to violate user's real assets.	**Skilled outsider**
W04: Identity thief.	**Skilled outsider**
W05: Hacking into user's account or/ and Facebook account.	**Skilled outsider**
W06: Stalking, people would follow users and their dogs according to their posts.	**Unskilled or intentional insider**
W07: Kidnapping, people would follow users and their dogs according to their posts.	**Unskilled or intentional insider**
W08: People give false instructions to the RTT system.	**Intentional insider**
W09: Direct attack to mongo database.	**Skilled outsider**
W10: Disputes in the RTT develop team.	**Skilled insider**
W11: DDOS-attack.	**Skilled outsider**

Table 4-1 Privacy threats list of RTT

4.3.3. Interview

Besides workshop, an interview is made as another means to collect data. Interview is a kind of conversation where one person ask questions and one person answers questions. The one asks questions is interviewer and the other one is interviewee. The questions are often designed in advance. Interview is a common method to understand people's opinions.

After the workshop, participants answered some interview questions. The interview questions are attached as appendix at the end of this thesis. Interview questions are mainly focused on participants' subjective feelings about the workshop and their

evaluation to their own work. A few points in these answers shall be highlighted, which might be helpful to analyze the result.

All participants are satisfied with the final result. Two of them give a 5 out of 5, and one gives a 4 out of 5. Participants think they find valuable privacy threats during this workshop. The second point is that, all participants indicate they have no idea whether they can get valuable findings in the workshop. Two participants said that the most difficult part is to start it at the beginning. That reveals the value of the DFD in this workshop. DFD is a description of the system. It contributes a lot to identify privacy threats. However, in this workshop, it is assigned as a task according to the workshop owner's knowledge. Is it the only option, or the best option to represent the system? It needs further discussions, which will be made in Chapter 5.

4.4. Procedure of LINDDUN step by step

4.4.1. Brief process of applying the LINDDUN methodology

As shown in Figure 3-1, there are six processes to apply the LINDDUN methodology to a system. The first step is to model a data flow diagram (DFD) of the system. Then, the analyst needs to locate all privacy threats associated with every data flow diagram element with the help of privacy threat types. The third step is to elicit and document privacy threats. LINDDUN has a strong support in this step, called threat tree. The analyst uses the threat tree as a check list to get specific privacy threats from last step. Once the analyst gets one specific privacy threat, he documents this privacy threat as a misuse case. A list of misuse cases is supposed to be the output of the third step. The forth step is to prioritize all privacy threats. Privacy threats with high risks are selected to be solved in next step. A mitigation strategy tree is given by the LINDDUN methodology as a strong support. It is used to help the analyst to find strategies for every privacy threat selected from last step. Finally, a privacy enhancing technology list offers solid techniques to match all strategies. According to these strategies and technologies, a list of privacy requirements can be produced in details.

4.4.2. Model data flow diagram of RTT

The importance of data flow diagram has been discussed in previous chapters. In this case study, the author created the data flow diagram of the RTT system according to the RTT project resources. Following analysis and discussions are all based on this data flow diagram of the RTT system, which is shown in Figure 4-4.

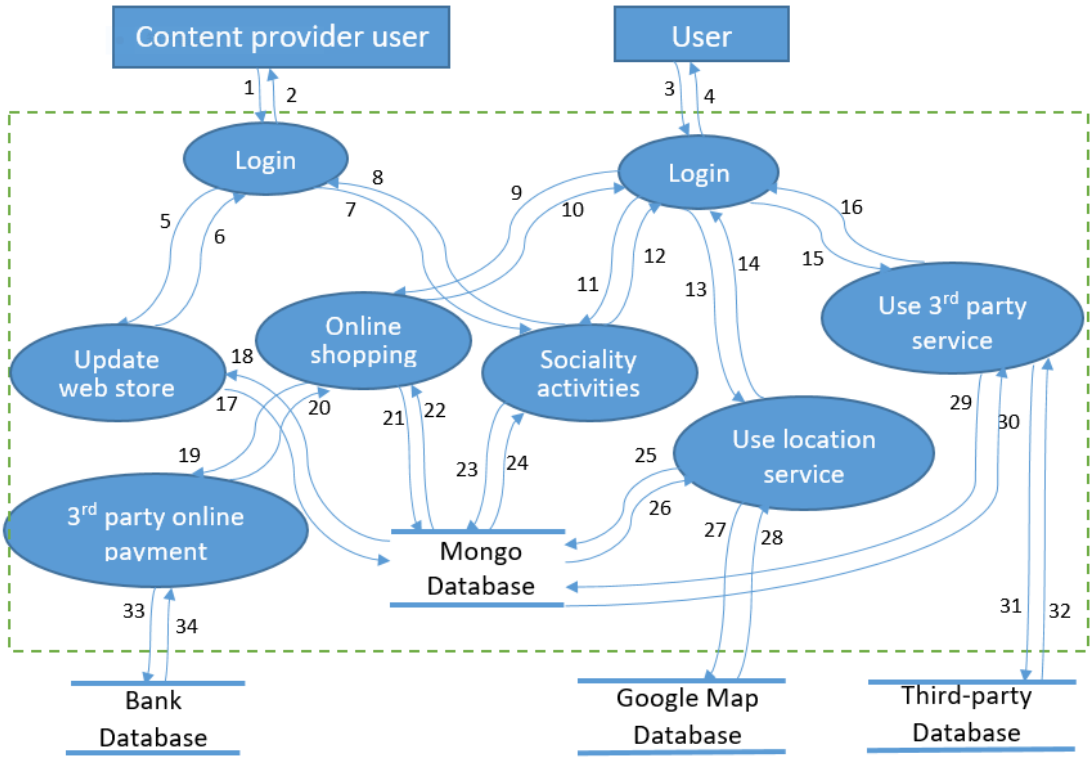


Figure 4-4 A data flow diagram of the RTT system

The data flow diagram of the RTT system contains 2 entities: the content provider user and the normal user. The content provider user is someone who updates the RTT website. The normal user is someone who get the RTT services. Besides, there are 8 processes, 4 data stores and 34 data flows. Limited to diagram size, descriptions of data flows are listed in Table 4-2. Some data flows in this table are users' PII. Some data flows, such as location information and payment information are PII which is created by the system. Besides, transaction data can be turned into PII, like import data request. So every data flow in this DFD shall be protected. This data flow diagram is based on data transferring in the RTT system. For example, when a normal user tries to use RTT to buy online dog training courses, he firstly logs in the RTT system. Through login process, he provides his identifier with data flow 3, and he gets permission with data flow 4 to use the RTT service. Then, he opens the RTT web store page to select what he needs. After that, he pays for the online course and gets what he bought. Data flow 9 is lookup and selection information from user, and data flow 19 is the information of user's credit card. Data flow 10 is the product information after this shopping action, and data flow 20 means a confirmation from payment system. On the other hand, data flow 33 and 34 are some transactions between the RTT system and bank system. One point needs to be highlighted is that login process is a special process in this data flow diagram. This is the first process for all users when they do communication with the RTT system, like a portal for user. It means that users need to get information from a certain process, which is impossible in a practical situation. Besides login process, data flow 1, 2, 3 and 4 might contain all information which other data flows contain.

Considered that content provider users are different from normal users, they have their own login process.

Data flow	Information description	Data flow	Information description
1	Identifier and almost all information this user provides to the RTT system.	2	Grant after confirmation and almost all information the RTT system provides to this user.
3	Identifier and almost all information this user provides to the RTT system.	4	Grant after confirmation and almost all information the RTT system provides to this user.
5	New web store updates information	6	Last web store version review
7	New PII	8	Other users' PII
9	Lookup and selections information	10	Products information
11	New PII	12	Other users' PII
13	Location information and usage permission	14	Location information
15	Import data permission (Authorization)	16	3rd party data confirmation
17	New web store updates information	18	Last web store version review
19	Credit card information	20	Payment confirmation
21	Shopping records	22	Updated products information
23	New PII	24	Other users' PII
25	Location records	26	Location records review
27	Location request	28	Location information
29	3rd party data records	30	3rd party data review
31	Import data request	32	3rd party data
33	Payment transaction information	34	Payment transaction information

Table 4-2 Data flow descriptions

There is a dashed box, which is not a part of data flow diagram in this data flow diagram. That is the boundary of the RTT system. The data stores out of this boundary are not belong to the RTT system and shall be omitted.

4.4.3. Map privacy threats to DFD elements

Mapping privacy threats to DFD elements means to create a table which contains all potential privacy threat points for every DFD element. The result of mapping privacy threats to DFD elements is shown in Table 4-3. The crosses in this table mean all potential privacy threat points for the RTT system so far. These points need to be further discussed in following steps.

	Threat target	L	I	N	D	D	U	N
Entity	Content provider user, e1	×	×				×	
	User , e2	×	×				×	
Data flow	Df1	×	×	×	×	×		×
	Df2	×	×	×	×	×		×
	Df3	×	×	×	×	×		×
	…	.	.	.	.	.		.

	Df34	×	×	×	×	×		×
Process	Login(c), p1	×	×	×	×	×		×
	Login(u), p2	×	×	×	×	×		×
	Update web store, p3	×	×	×	×	×		×
	Online shopping , p4	×	×	×	×	×		×
	Sociality activities, p5	×	×	×	×	×		×
	Use location service, p6	×	×	×	×	×		×
	Use 3rd party service, p7	×	×	×	×	×		×
	3rd party online payment, p8	×	×	×	×	×		×
Data store	Mongo database, ds1	×	×	×	×	×		×
	Bank database, ds2	×	×	×	×	×		×
	Google map database, ds3	×	×	×	×	×		×
	Third-party database, ds4	×	×	×	×	×		×

Table 4-3 Map LINDDUN to threats DFD elements of RTT

Making assumptions is helpful to reduce the number of potential privacy threat points. It is greatly influenced by the system features. With the help of the LINDDUN official tutorial [Wuyts and Joosen, 2015] and the use case diagram of the RTT system, assumptions in Table 4-4 are made. There are some assumptions are made directly according to the LINDDUN tutorial. A1, A2 and A3 are stated here, because that the 'X's is the same when it involves the same type of data. A6, A7, A8, A9, A10 and A11 are optional assumptions to general systems. Linkability and identifiability are only applicable to anonymous systems, and non-repudiation and detectability are only applicable to the systems with e-voting or whistleblowing activities. A12 is based on the rule that all 'X's related to processes can usually be determined as not applicable to the system. A13 and A15 are directly recorded in the LINDDUN tutorial and they are applicable to all systems. All assumptions above are based on the LINDDUN tutorial. The rest assumptions: A4, A5 and A14 are made by the author's analysis to the RTT system. Some assumptions are called general assumptions, such as A1, A2, A3, A4, A9, A12, A13 and A15, as they are applicable for most systems. The rest assumptions are only applicable for certain systems. The suggestions on how to make assumption are unclear and not summarized well either in tutorial or in Wuyts' thesis. The analyst might be confused. This is a point which can be improved in LINDDUN. It is valuable to further discuss it in chapter 5.

A1. All internal processes are processing in a similar way. It can be assumed that the RTT system is mainly threaten by outside threats, all internal processes can be handled as one.

A2. All data flows between internal processes, and between internal processes and internal data stores, are processing in a similar way. It can be assumed that the RTT system is mainly threaten by outside threats. All internal data flows can be handled as one.

A3. All data flows between internal processes and external data stores are processing in a similar way. All these data flows can be handled as one.

A4. Data flows between an entity and a process, and between process and data store are

not trusted
A5. Internal data stores are not considered confidential. Mongo database, as a database service, shall not be trusted.
A6. Linkability and identifiability of content seller is not applicable. The information of the company public account is necessary to be real and public.
A7. Linkability and identifiability of user can be susceptible if users use nick name instead of real identity, or the user want to hide part of their information. The RTT system can be anonymous. Linkability and identifiability of user might be a threat in the RTT system.
A8. Linkability and identifiability of content seller to portal data flow is not applicable, because of assumption 7.
A9. As linkability and identifiability of data flows and processes have similar threats and solutions in the RTT system, they can be handled as one. A distinction between these two threats of data flows and processes will therefore not be made.
A10. Non-repudiation is not applicable in the RTT system. There is no e-voting or whistleblowing activities in the RTT system. So as a social network application, users shall not deny their usage of the RTT service.
A11. Detectability is not applicable in the RTT system. For a social network web application, users are not supposed to deny their usage of the RTT service. However, it is possible for users to use other third party services. The data flow between the RTT system and third-party data stores, like payment transaction, have to be considered as a detectability threat. The solution to this problem shall be connected to information disclosure of data flow.
A12. Disclosure of information is not applicable to internal processes. Internal processes are assumed protected well. As the same reason, it is not applicable to data flows within internal processes.
A13. Disclosure of information is not a privacy threat in LINDDUN, but a security threat in STRIDE. So when it comes to disclosure of information threats, the corresponding security threat trees will be used instead of privacy threat trees.
A14. Content unawareness of two entities are pretty similar on problem and solution. A distinction between the two entities will therefore not be made.
A15. Non-compliance is an important threat, however, it is not specific to one part of the system, but poses to the system as a whole. A distinction between the different DFD elements for this threat will therefore not be made.

Table 4-4 Assumptions of the RTT system

According to these assumptions, meaningless potential privacy threat points are removed, and similar potential privacy threat points are combined. A new privacy threats table is produced as shown in Table 4-5. There are totally 11 potential privacy threat points in this table. Each number stands for one point which shall be noticed in threats eliciting. This LINDDUN threat table presents a warning list to LINDDUN analysts. It is not supposed to find only one privacy threat for every number. Each threat

might be connected to several point numbers. On the contrary, every point number might be related to plural threats, or even none. These assumptions shall be fixed after this table is done. However, if there is something unclear in following steps, more assumptions shall be added to make things clear and definite.

	Threat target	L	I	N	D	D	U	N
Entity	Content provider user, e1						3	
	User , e2	1	2				3	
Data flow	Df1, 2					4		11
	Df3, 4	5	5			4		11
	Internal data flows between processes (df5-df16, df19 and df20)							11
	Data flows between process and internal data store (df17, df18, df21-df26, df29 and df30)	6	6			4		11
	Data flows between process and external data stores (df27, df28 and df31-df34)	6	6		7	4		11
Process	Internal processes (all)					8		11
Data store	Mongo database, ds1	9	9			10		11

Table 4-5 Simplified LINDDUN threat table after combine "X"s

4.4.4. Elicit and document threats

On the basis of potential privacy threat points list, next step is to elicit and document privacy threats with threat tree. Threat tree is one contribution given by LINDDUN to help analyst to elicit privacy threats from potential privacy threat points. It is live updated online[11] [Wuyts et al., 2015]. Misuse case template is used as the template to document threats.

The process to elicit privacy threats is introduced in previous chapters. According to last step's result, there are 11 potential privacy threat points and every point is related to one or two privacy threat tree. First potential privacy threat point is taken as an example to show how these privacy threats are elicited and documented from potential privacy threat points with the help of privacy threat tree.

A privacy threat tree shall be used as a checklist. First potential privacy threat point is linkability of normal user. Its corresponding privacy threat tree is shown in Figure 3-3. In Figure 3-3, "linkable login using untrusted communication (L_e1)" is a possible

[11] https://distrinet.cs.kuleuven.be/software/linddun/catalog.php

problem for the RTT social network. Its leaf nodes: linkable login and untrusted communication, shall be both privacy threats to the RTT system. Log in is a basic feature for the RTT social network system in the RTT use case diagram. A normal user of the RTT service might use fixed username and password combination for several social network accounts, and once one of them leaks, all other account are in danger. Therefore, first threat shall be login threat, and further details shall be discussed according to the RTT system implementation. Secondly, untrusted communication can be caused by different reasons. Besides information disclosure of data flow and linkability of data store, another one is that the information receiver is not trustable. In this case study, the RTT system needs to collect users' information for the RTT service, such as dog's name, age, brand, some pictures, and so on. The information receiver is the company who runs the RTT services. If they are not trustworthy, the information is not secured. Then, second privacy threat is untrustworthy information receiver threat. Review these two leaf nodes with more details, then two privacy threat are elicited, and documented as shown in Table 4-6 and Table 4-7.

Finally, 15 privacy threats are elicited and documented, and they are attached in Appendix 1. The problem-oriented step, as described in previous chapters, is the core of LINDDUN. Up to this step, all problem-oriented procedures have been finished.

Threat 01. Login threat.

Summary	Users' login information (account ID, password, profile…) is linkable to other logins, or too specific which could be an identity, or too much information contained in login process. The account will be linkable to other services.
Primary misactor	User.
Basic flow	Bf1: Users use fixed information as id (or password, profile…) to login. Bf2: Potential attackers find similar ID or related information, then get more information which should not be open. Consequence: Potential attackers get more misactors' information than expected.
*Leaf node(s)**	L_e2, L_e4, L_e5
*Root node(s)**	Linkability of entity, identifiability of entity
*DFD element(s)**	Normal user

Table 4-6 Threat 0. Login threat.

Threat 02. Untrustworthy information receiver threat..

Summary	Information receivers (service provider), or a part of users are not trustable.

Primary misactor	User, information receiver.
Basic flow	Bf1: Users provide information. Bf2: The information is disclosed by misactor. Consequence: Potential attackers get users' information through misactor.
*Leaf node(s)**	L_e3, L_e6, I_e3, I_e6
*Root node(s)**	Linkability of entity, identifiability of entity, non-compliance.
*DFD element(s)**	Normal user

Table 4-7 Threat02. Untrustworthy information receiver threat

4.4.5. Threats prioritization

After three problem-oriented steps, LINDDUN elicits 15 threats from the RTT system. Next step is to make prioritization. The OWASP Risk Rating Methodology (ORRM) [2015] is chosen as a method to estimate the severity of all of these threats in this thesis.

Step 1. Identifying a Risk: After first three steps, 15 threats are elicited from the RTT system, as shown in Table 4-8.

Threats
L01: Login threat.
L02: Untrustworthy information receiver threat.
L03: Too much information sharing threat.
L04: Unawareness storage of information.
L05: Spoofing a user of the RTT system by falsifying credentials.
L06: Spoofing a user of the RTT system by eavesdropping communication.
L07: Credential is disclosed.
L08: Communication session token is disclosed.
L09: Information disclosure of internal process.
L10: Communication content is disclosed.
L11: Communication with third party is not protected.
L12: Third party is untrustworthy.
L13: Database is untrustworthy.
L14: Incorrect or insufficient privacy policies.
L15: Non-compliance insider actions.

Table 4-8 Privacy threats list for the RTT system

Threats	Skill level	Motive	Oppor-tunity	Size	Ease of discovery	Ease of exploit	Aware-ness	Intrusion detection	Average

L01	1	1	9	9	7	5	4	8	5.5
L02	1	4	0	2	3	9	1	9	3.625
L03	1	1	9	9	9	9	1	8	6.875
L04	1	1	9	9	3	5	4	3	4.325
L05	9	9	7	2	3	5	6	3	5.5
L06	9	9	7	2	3	5	6	3	5.5
L07	5	9	4	2	3	5	4	3	4.325
L08	9	9	4	2	3	5	4	1	4.625
L09	3	4	4	2	3	5	1	1	2.875
L10	9	9	4	2	3	5	4	1	4.625
L11	9	9	4	2	3	5	4	1	4.625
L12	1	4	0	2	3	9	1	9	3.625
L13	9	9	7	2	3	9	6	3	6
L14	5	4	9	4	3	3	4	1	4.125
L15	3	4	4	2	1	9	1	1	3.125

Table 4-9 The likelihood point of threats

Step 2. Factors for Estimating Likelihood: In order to estimate likelihood, a series of factors shall be taken into consideration. In OWASP Risk Rating Methodology (ORRM), every factor is influenced by a set of options, and ORRM rate these options from 0 to 9. Eight factors in ORRM are taken into consideration in this case study: skill level, motive, opportunity, size, ease of discovery, ease of exploit, awareness, intrusion detection. According to ORRM's rating descriptions [OWASP, 2015], every threat get an average point of likelihood, as shown in following Table 4-9.

Step 3. Factors for Estimating Impact: According to OWASP Risk Rating Methodology, in order to estimate impact, a series of factors shall be taken into consideration. Every factor is influenced by a set of options, and in ORRM, these options are rated from 0 to 9. Social network is kindly different from normal information system, and privacy threats risking is not totally the same with security threats, so following 4 factors are selected in our model: loss of confidentiality, loss of accountability, reputation, privacy violation size. According to ORRM online guides, every threat can get a final average point of impact, as shown in following table.

Threats	Loss of confidentiality	Loss of accountability	Reputation damage	Privacy violation size	Average
L01	9	9	1	3	5.5
L02	9	1	9	9	7
L03	7	9	1	5	5.5
L04	6	9	4	7	6.5
L05	9	7	5	5	6.5

L06	9	7	5	5	6.5
L07	9	7	4	7	6.75
L08	2	7	4	5	4.5
L09	6	7	5	3	5.25
L10	2	7	4	5	4.5
L11	2	7	1	7	4.25
L12	9	1	4	9	5.75
L13	9	7	9	9	8.5
L14	2	7	9	5	5.75
L15	6	1	9	7	5.75

Table 4-10 The impact point of threats

Step 4. Determining Severity of the Risk: According to point table in ORRM, point in (0, 3] can be rated LOW, point in (3, 6] can be rated MEDIUM, and point in (6, 9] can be rated HIGH, as shown in Table 3-3. Combine likelihood and impact table, final severity has five ranks: CRITICAL, HIGH, MEDIUM, LOW and NOTE, as shown in Table 3-4. According to the points of every privacy threat, the ranked threat list is shown in Table 4-11.

Threats	Likelihood	Impact	Overall Severity
L13: Database is untrustworthy.	HIGH	HIGH	CRITICAL
L02: Untrustworthy information receiver threat	MEDIUN	HIGH	HIGH
L03: Too much information sharing threat	HIGH	MEDIUN	HIGH
L04: Unawareness storage of information	MEDIUN	HIGH	HIGH
L05: Spoofing a user of the RTT system by falsifying credentials.	MEDIUN	HIGH	HIGH
L06: Spoofing a user of the RTT system by eavesdropping communication.	MEDIUN	HIGH	HIGH
L07: Credential is disclosed.	MEDIUN	HIGH	HIGH
L01: Login threat	MEDIUM	MEDIUN	MEDIUM
L08: Communication session token is disclosed.	MEDIUN	MEDIUN	MEDIUM
L10: Communication content is disclosed.	MEDIUN	MEDIUN	MEDIUM
L11: Communication with third party is not protected.	MEDIUN	MEDIUN	MEDIUM
L12: Third party is untrustworthy.	MEDIUN	MEDIUN	MEDIUM
L14: Incorrect or insufficient privacy policies	MEDIUN	MEDIUN	MEDIUM
L15: Non-compliance insider actions	MEDIUN	MEDIUN	MEDIUM

L09: Information disclosure of internal process.	LOW	MEDIUN	LOW

Table 4-11 The overall severity of threats

Step 5. Deciding What to Fix: In this case study, only threats which are rated HIGH or CRITICAL are selected to be solved. Finally, first seven threats in Table 4-10 are decided to be analyzed in further analysis.

Step 6. Customizing Your Risk Rating Model: Risk rating model is not a study objective in this thesis. The result produced by previous work is enough to continue following analysis in this case study. Customizing model is not discussed.

4.4.6. Elicit mitigation strategies

After threats prioritization, 15 privacy threats are elicited from the RTT system, and 7 of them are selected as targets to be focused. They are L02, L03, L04, L05, L06, L07 and L13. Finally, a privacy requirements list, in which there are 7 privacy requirements, is produced to solve selected privacy threats. L13 is taken as an example here. The root nodes of this threat are L_ds, I_ds and ID_ds. According to Table 3-5, the mitigation strategy to information disclosure of data store is confidentiality in guard exposure catalog. Then, from Figure 3-4, the system shall use access control and encryption to ensure the security of the data store, which is the first requirement R01 in the requirements list. This privacy requirements list almost covers all privacy mitigation strategies in the taxonomy of privacy mitigation strategies. Due to the universality of the RTT system, this result is also valuable to other social network sites or applications. The list in Figure 4-12 is one important contribution of this thesis. What's more, the final step of LINDDUN: Selecting privacy enhancing technologies, is not the focuses of this thesis, and is not discussed.

Requirement ID	Threat ID	Requirement definition
R01	L13, L07	The system shall have access control or encryption for database protection.
R02	L13	The system shall minimize user data on the database.
R03	L02, L03	The system shall protect user's ID by pseudonyms or other means.
R04	L03, L04	The system shall have feedback awareness tools; The system shall have friendly privacy support as default setting.
R05	L04	User data which is collected and used by the system shall be clear and accessible to user.
R06	L05, L06	The system shall hide user data when it stores or transfers user data, and shall remove the data once the data is

		outdated.
R07	L05, L06	The system shall protect transactional data by strictly obeying protocols and rules in transactions, or removing the data once the data is outdated.

Table 4-12 Privacy requirements list of the RTT system

5. Discussions

5.1. Verification of the results

The LINDDUN methodology and the workshop produced two privacy threats lists. List L is from the LINDDUN methodology and list W is from the workshop. The correctness and completeness of list L and list W shall be discussed in this section.

Limited to the author's knowledge, it is hard to verify the correctness of all threats in two lists. But there is a way to narrow the number of threats which shall be focused. The Microsoft privacy guideline is used as a reference in this verification. If there is at least one rule or suggestion, which solves the threat, this threat is regarded as a valuable one. The threats which are not mentioned in the Microsoft privacy guideline need further discussions. As shown in Table 5-1 and Table 5-2, 10 threats in list L and 7 threats in list W can get corresponding guides in the Microsoft privacy guideline. These 17 are considered correct privacy threats to the RTT system.

Threats in list L	Guides in the Microsoft privacy guideline
L01	Scenario3. Must provide user with prominent notice, and get explicit consent prior to collection. Should be provided in the UI, not license agreement.
L03	Scenario1. Should use data validation controls to filter out inconsistent, incomplete or incorrect PII. Scenario6. Must store minimum amount of data, for the shortest amount of time necessary to achieve business purpose.
L04	Scenario6. Must provide a secure mechanism for users to access and correct stored PII.
L05,06,07	Scenario6. Must authenticate users via a company-approved process before collecting, displaying, or modifying PII or contact preferences. Must store PII using appropriate security mechanisms to help prevent unauthorized access. Must restrict PII access to those with a need to know, and revoke access when no longer needed.
L11,12	Scenario7. Provide separate explicit Opt-In consent mechanism. Provide link to third party privacy statement.
L13	Scenario7. Discoverable Notice is required.

Table 5-1 Corresponding guides of List L

Threats in List W	Guides in the Microsoft privacy guideline
W01	Scenario1. Should not use methods of form submission that potentially expose data in a web form intended for or likely to result in the collection of PII. Must provide prominent notice and get explicit opt-in consent at any point prior to transfer.
W03,06,07	Scenario7. Provide separate explicit Opt-In consent mechanism. Scenario6. Must store minimum amount of data, for the shortest amount of time necessary to achieve business purpose. Must provide a secure mechanism for users to access and correct stored PII.
W04, 05	Scenario6. Must authenticate users via a company-approved process before collecting, displaying, or modifying PII or contact preferences. Must store PII using appropriate security mechanisms to help prevent unauthorized access. Must restrict PII access to those with a need to know, and revoke access when no longer needed.
W09	Scenario7. Discoverable Notice is required.

Table 5-2 Corresponding guides of List W

The rest threats in list L are L02, L08, L09, L10, L14 and L15, and the rest threats in list W are W02, W08, W10 and W11. These threats need be analyzed with a comparison. A comparison table is shown in Table 5-3.

ID	Related threats in list L	ID	Related threats in list W
W01	L01	**L01**	W01
W02	X1	**L02**	X5
W03	L03, 04	**L03**	W03, 06, 07
W04	L05, 06	**L04**	W03, 06, 07
W05	L07	**L05**	W04
W06	L03, 04	**L06**	W04
W07	L03, 04	**L07**	W05
W08	X1	**L08**	X2
W09	L13	**L09**	X2
W10	X05	**L10**	X2
W11	X0	**L11**	X3
		L12	X3
		L13	W09
		L14	X4
		L15	X5

Table 5-3 Comparison between list L and list W

This table shows a connection between list L and list W. As we can see, almost all threats in list W have at least one correspongding threat in list L, except for W02, W08 and W11. Similarly, L08, L09, L10, L11, L12 and L14 have no connections to list W.

Firstly, W11 which is marked X0 in the table shall be removed. Because a distributed denial-of-service attack (DDoS) is a security threat, and is far to privacy disclosure. W02 and W08, which are marked as X1, are caused by user's unintentional actions. These actions might lead to privacy disclosure, and deserve to be a privacy threat. W02 and W08 have no connection with list L, which means this privacy threat might be missed by the LINDDUN methodology. X2 means threats related to transactions inside a system, including L08, L09, L10. These threats are security threats, but can lead to a privacy disclosure. They are valuable privacy threats. L11 and L12 are both threats related to third-party. They are mentioned in the Microsoft privacy guideline as valuable privacy threats. They are marked X3 as a missing point of list W. L14 is a policy related threat. L14 is not mentioned in the Microsoft privacy guideline, but "notice and consent" appears in almost every other threat. L14 is marked X4 as another missing point of list W. X5 shows in both lists. It marks the threats which have no connections to the Microsoft privacy guideline. W10, L02 and L15 are threats caused by insiders' intentional actions. It shall be removed, because for a social network website like RTT, all group members are trustable. X1, X2 and X3 are three types of privacy threats which shall be focused after the verification.

THREAT CATEGORIES	Entity	Data Flow	Data Store	Process
Linkability	×	×	×	×
Identifiability	×	×	×	×
Non-repudiation		×	×	×
Detectability		×	×	×
Information Disclosure		×	×	×
Content Unawareness	×			
Policy/consent noncompliance	√	×	×	×

Table 5-4 Modified privacy threats map table

X1 threat is one meaningful threat in list W. There are two threats marked by X1: W02 and W08. They are both caused by user's unintentioanl actions. People might give false instructions unintentionally to social network applications. Somethimes, it is harmful to uesr's privacy. The users shall be resposible for this privacy threat. However, if there are solutions to solve this privacy threat, engineers and analyst should do so to make their systems more reliable. Privacy property of this threat shall be Non-compliance of Entity. This privacy threat type is ignored at the beginning. Non-compliance of entity is not included in original the LINDDUN potential privacy threat list. Due to the incompleteness, non-compliance of entity is regarded as one flaw of the LINDDUN framework. A new mark on the privacy threats map means to a series work. A new

threat tree related to non-compliance of entity is necessary if this new threat is meaningful to the LINDDUN methodology. The constuct of non-compliance of entity might be a future work of this thesis.

X1 is the weak point of the LINDDUN methodology, while X2 and X3 reflect the advantages of the LINDDUN methodology. Privacy threats marked by X2 are all related to transaction data, which refers to internal activities of a system. These activities are easily ignored by developers and engineers in privacy requriement analysis. For the LINDDUN methodology, all processes and data flows are expressed in the DFD at the beginning. Privacy threats marked by X3 are the threats caused by third parties or agents. It is another blind spot for some requirements elicitation techniques like workshop. Additionally, there are 9 valuable privacy threats out of 11 in list W, from W01 to W09. In list L, 13 out of 15 privacy threats are considered valuable, and six of them are missed by the workshop. This result matches the first hypothesiss before the case study: the LINDDUN methodology provides a systematic way to identify privacy threats, and processes more privacy threats than workshop. For the second hypothesis, only 9/13 of privacy threats match the Microsoft privacy guideline. It is less than we expect.

5.2. Discussion on data flow diagram

Data flow diagram is important as the first step of LINDDUN. It is the base of latter procedures, and it has great influence on the final result. When the author applies the LINDDUN methodology to the RTT system, there are two questions at the beginning.:

1. How much information shall be included in the data flow diagram of the system?
2. Is data flow diagram the only option for the LINDDUN methodology?

There are two findings from the case study. Firstly, processes and data flows of the RTT system can be simplified. In this case study, the author has ever changed the data flow diagram in step 2 and step 3. For example, in the use case diagram of the RTT system, location feature is combined by two use cases: check in on map and get location. These two use cases could be two separate processes in the data flow diagram. As both of them are dealling with location information and they are both internal processes in system, these two processes are combined as one. This modification did not matter the final result. It indicated that some processes could be simplified when a data flow diagram is created. For the example above, use location service process is enough for the latter analysis, instead of check in on map process and get location process. The details within a function model can be simplified as one process. There is a general assumption in the LINDDUN tutorial for social websites matches this conjecture: for a social website, *”Internal DFD elements are considered trustworthy.”* [Wuyts and

Joosen, 2015]. In the RTT system, all processes and data flows within processes are internal DFD elements. They are all considered trustworthy, and removed in following procedures. Secondly, some data flows can be combined, but others cannot. There are two data flow diagrams respectively shown in Figure 4-3 and Figure 4-4. Besides the differences between lines and arrows, latter has 8 more data flows, such as df_17 and df_18. These 8 data flows are considered same as df_21 and df_22, and combined as one. The difference is mentioned and removed by assumption A2. However, some data flows, such as df_1 and df_2, cannot be removed, and they lead to potetial privacy threat points in Table 4-5. The two findings above reveal that assumptions are helpful to simplify a DFD. According to the general assumption for social websites: *"Internal DFD elements are considered trustworthy"*, processes and data flows within processes can be abandon in later analysis. In another word, internal processes and data flows within processes could be simplified or removed. Data flows which are connected to entities and data stores deserve more attetion.

In order to identify all hidden privacy threats and elicit all privacy requirements, analysts will try to present every detail of the system in a DFD. However, a simplified DFD with necessary processes and data flows might reach the same goal. Duplicate actions and redundant content could be simplified to make the DFD clear. Data flow diagram is working as a reminder to remind analyst where the weakness of the system is. Instead of covering every specific process of the system, it is more important to draw attetion on the weak points of the system.

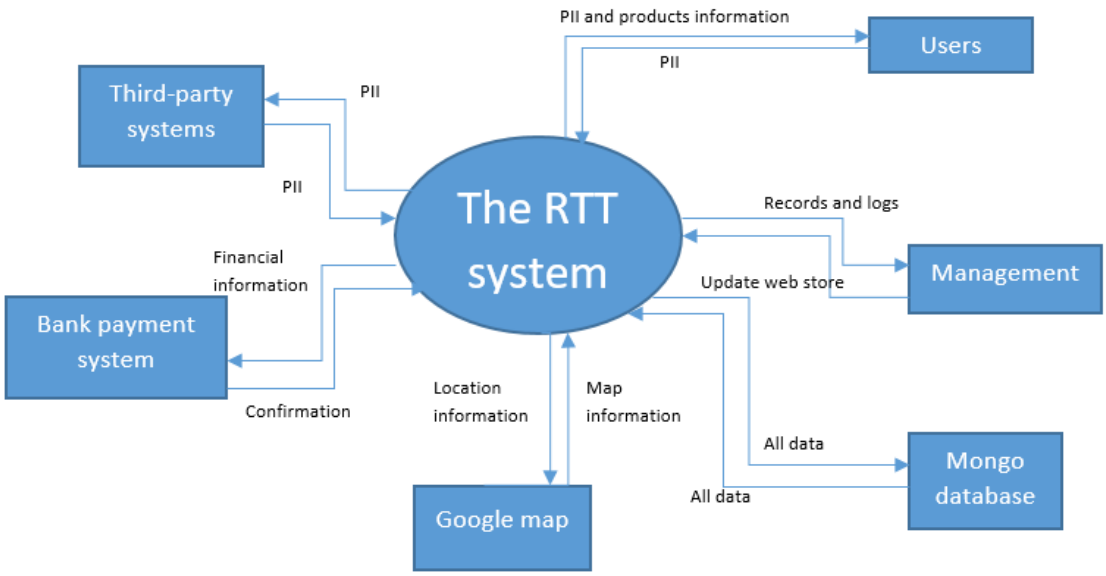


Figure 5-1Context diagram of the RTT system

The second question is about other options than DFD. Figure 5-1 presents a simple context diagram of the RTT system. It contains all external entities related to RTT

system. Compared to the data flow diagram in Figure 4-4, there are many similarities between these two diagrams. Firstly, they both express some activities in the RTT system. Secondly, they both shows data types which the system uses to communicate with external entities and systems. These two similarities are both needed by the LINDDUN methodology. There is one difference between these two diagrams. The data flow diagram has a clear view of inside processes and data flows, while the context diagram does not. In Table 4-5, all potential privacy threat points related to internal data flows and internal processes are removed according to assumptions, except for the detectability of process. The detectability of process does not have an independent threat tree in the LINDDUN privacy threat tree catalog. It is refered to the information disclosure of process tree, which is assumed protected well in the RTT system in assumption A12. This result matches the general assumption in the LINDDUN tutorial again, *"Internal DFD elements are considered trustworthy"*. In conclosion, at least for the RTT system, context diagram is able to take the place of the DFD without great influence.

The general assumption in the LINDDUN tutorial *"Internal DFD elements are considered trustworthy"* is for social websites. It still needs more studies and discussions when context diagram is applied to other system types. For example, a CS model game system contains two part: client and server. Both sides need protection on users' information, as well as the communication between them. A context diagram cannot show these communications in a good view, because they are inside the system. In this case, a DFD is better to be applied. However, it shows another way to improve the LINDDUN methodology: start analysis with another a chosen diagram. A good expection is that the LINDDUN methodology keeps two or more different diagrams at the same time. For some systems which are simple and clear, like the RTT system, context diagram might be better. For other systems with more internal details, data flow diagram is a better choice which shows a clear internal view. A new system diagram type needs proper assumptions and privacy threat tree catalog. These are the future work of this thesis.

5.3. Discussion on making assumptions

When the author uses LINDDUN as an analyst, another difficult proccedure is to make assumptions. Making assumptions is an important procedure and deserves more instructions because of its importance. Every assumption results in adding or removing privacy threats from the privacy threats list. Secondly, making assumptions can fix problems of the data flow diagram, which is discussed in chapter 5.2, and make the LINDDUN custom. What's more, it is a necessary process to reduce the workload.

No.	System description	Assumptions
1	General assumptions	Combine elements which involve the same data type and apply to the same threat.
2		Non-compliance threats should not be applied to a specific DFD element, but are applicable to the entire system.
3	Social network	Internal DFD elements are considered trustworthy.
4		Non-repudiation and detectability threats are considered irrelevant.
5	The systems without anonymous credentials, anonymous communication	Linkability and identifiabilit of entity are not applicable.
6	The systems without e-voting, whistleblowing	Non-repudiation of data store and data flow is not applicable.
7	The systems that do not handle highly sensitive data.	Threat types of processes are often not applicable or low priority.
…	…	…

Table 5-5 A brief guide to make assumptions

LINDDUN is growing and updated aperiodically by its research group. The version of the LINDDUN methodology in this thesis is July, 2015. Only limited resources in the LINDDUN official tutorial are helpful to make assumptions. The researchers on the LINDDUN methodology have made much effort to improve assumptions hints, but they are still not clear enough. In the LINDDUN official tutorial, there are some tips on making assumptions: combining ’X’s and general assumptions [Wuyts and Joosen, 2015, p27]. If several ’X’s involve the same type of data, they can be combined when they apply to the same threat. All non-compliance ’X’s shall be combined and applied to the entire system. These two rules are applicable for most systems. Besides, there are still 5 rules only applicable for specific systems. Internal DFD elements are considered trustworthy for social networks. The 4th rule removes non-repudiation and detectability for social networks. The 5th rule states that linkability and identifiability are only applicable to the systems which support anonymous use. The 6th rule indicates the situations which non-repudiation of data store and data flow are applicable. Finally, the 7th rule points out one situation which threat types of processes are applicable. All these rules which appears in the LINDDUN tutorial can be summarized as Table 5-5. the tab system description shows the conditions when the assumptions are applicable. And the assumptions tab are all summarized by experience of LINDDUN’s authors. Enhancing guides to make assumptions is a good direction to improve LINDDUN. A brief rules list

to make assumptions shall be helpful. Considering there are many check list in the LINDDUN methodology, this rules list can be a new check list to help analysts make their own assumptions. This list can be extended when new rules are found by researchers of LINDDUN.

6. Conclusion

In this thesis, literature related to privacy concepts has been reviewed. Privacy issues gains increasing attention in now days. Proper tools and methods to deal with privacy issues shall be created and improved to meet people's privacy requirements. LINDDUN is one of them. The LINDDUM methodology is reviewed in this thesis. A case study on the LINDDUN methodology and a workshop about privacy threats elicitation are conducted. Discusssions on creating a data flow diagram, making assumptions and other factors are given and summarized. A few issues are highlighted: A. The LINDDUN methodology is a more systematic methodology compared to other requirements elicitaition techniques, like workshop. Using the LINDDUN methodology reduces the chance to miss privacy threats of a system. However, non-compliance of entity, which caused by people's unintentional actions, is missed by the LINDDUN framework as a flaw; B. Adding new diagrams, like context diagram, might be a good direction to rich the scope of application for the LINDDUN methodology; C. Hints and tips about making assumptions shall be more complete and specific. A simple summary of current hints and tips are organized as a guide list in section 5.3. An official guide on making assumptions will improve the usability of LINDDUN. These views will be helpful for LINDDUN's further improvement.

This thesis has several contributions. Firstly, one case study and one workshop are included and compared in this thesis. The discussion between two results of them implies some advantages and disadvantages of the LINDDUN methodology. Compared to workshop, the LINDDUN methodology lead the analyst to identify more privacy threats and get more privacy requirements. The procedure of LINDDUN makes analyzing process predictable, and makes the result more reliable. However, the LINDDUN methodology has a blind spot on non-compliance of entity. It ignores users' unintentional misactions. Secondly, this thesis discussed details of DFD and other possibilities instead of DFD. DFD is supposed to be as complete as possible to identify all possible privacy threats. Meanwhile, context diagram is appliable and can produce similar result compared to DFD for the RTT system. New diagrams will rich the scope of applications of LINDDUN. It is another direction to improve the LINDDUN methodology. Finally, a guide on making assumptions are summarized with the help of the LINDDUN official tutorial and given examples. As an important procedure in the LINDDUN methodology, assumption making deserves more cretirions.

There are some limitations in this thesis. The resources of LINDDUN are limited to one research group. The author lack means to compare the method with other privacy analysis methods. Second limitation is that the RTT system affects the result because of

its specifity. The RTT system only has basic social network features. Group members of the RTT team are all students. The workshop result is limited by participants' knowledge, which might be biased. What's more, the author is short of means to verify the result of neither the case study or the workshop. The verification is also a hard point for the LINDDUN methodology. The lack of verification shall be considered not only a limitation, but also a direction to do improvements.

LINDDUN is not the end of people's exploration to privacy issues. On the one hand, there is still space for LINDDUN to be practical and sufficient. There are many directions to improve LINDDUN. The concept framework could be more consummate, the procedures could have more optional factors, and even result verification is one improvement to LINDDUN. On the other hand, the LINDDUN framework is a good example of analyzing privacy requirements. Thanks to previous researchers, privacy issues is becoming more predictable and controllable in software development.

References

[Warren and Brandeis, 1890] Samuel Warren and Louis Brandeis, The Right to Privacy. *Harvard Law Review*, 4(193), 1890.

[Rachels, 1975] Rachels James, Why privacy is important? *Philos Public Aff*, 1975, 323–333.

[Cohen, 2000] Cohen, J., Examined lives: informational privacy and the subject as object. *Stanford Law Review*, 52(5), 2000, 1373-1438.

[Regan, 2002] Regan, P.M., Privacy as a common good. Information, *Communication and Society*, 5(3, 2002), 382-405.

[Gellman, 1998]Gellman R, Does Privacy Law Work? *Technology and Privacy: The New Landscape*, 1998, 193-218.

[Onn et al., 2005] Yael Onn et al., Privacy in the Digital Environment. *Haifa Center of Law & Technology*, 2005.

[Steeves, 2009] Steeves, V., Reclaiming the social value of privacy, *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked World*, 2009, 191-208.

[ISO 29101, 2013] ISO 29101, Information technology - Identification of privacy protection requirements pertaining to learning, education and training (LET), 2013.

[ISO 29100, 2011] ISO 29100, Information technology - Security techniques - Privacy framework, 2011.

[European data protection law, 2014] Council of Europe, *Handbook on European data protection law*, European Union Agency for Fundamental Rights, April, 2014.

[Federal Trade Commission, 2014] Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers, *Federal Trade Commission*, 43, 2014.

[Dennedy et al., 2014] Dennedy, Michelle, Fox, Jonathan, Finneran, Tom, *The Privacy Engineer's Manifesto*, Getting from Policy to Code to QA to Value, 2014.

[McCormick and Michelle, 2011] McCormick, Michelle. New Privacy Legislation. *Beyond Numbers*, 2011.

[Wiegers, 2003] Karl E. Wiegers, Software Requirements, 2003.

[Li and Chen, 2009] Qing Li and Yu-Liu Chen, *Modeling and Analysis of Enterprise and Information Systems*, 2009, 85-97.

[Deng et al., 2010] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, Wouter Joosen, A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements, 2010.

[Microsoft, 2002] The STRIDE Threat Model, Microsoft, Commerce Server, 2002, available as: https://msdn.microsoft.com/en-us/library/ee823878(v=cs.20).aspx

[Stan and Ewa, 2010] Digital Natives and Mobile Phones: A Survey of Practices and Attitudes about Privacy and Security. *2010 IEEE International Symposium on Technology and Society*, 2010.

[Danezis, 2008] G. Danezis, Talk: an introduction to u-prove privacy protection technology, and its role in the identity metasystem – what future for privacy technology, 2008.

[Guttorm and Andreas 2001] Guttorm S. and Andreas L., Templates for misuse case description, 2001, 1–13.

[Microsoft, 2008] Privacy in software development – secure software mad easier. Technical report, Microsoft Coorporation, September, 2008: available as http://download.microsoft.com/download/9/3/5/935520EC-D9E2-413E-BEA7-0B865A79B18C/Privacy%20in%20Software%20Development.ppsx

[Beckers, et al., 2014] Kristian Beckers, Stephan Faßbender, Maritta Heisel, and Rene Meis. A Problem-based Approach for Computer Aided Privacy Threat Identification. *Privacy Technologies and Policy*, **8319 of LNCS**, 2014, 1-16.

[Wuyts, 2015] Unpublished manuscript, Kim Wuyts, January 2015: available as https://distrinet.cs.kuleuven.be/software/linddun/index.php.

[Wuyts, 2015] Kim Wuyts, Privacy Threats in Software Architectures. PhD thesis, Ku Leuven, January 2015.

[Wuyts, 2015] Kim Wuyts, Privacy threat trees catalog. Available as: https://distrinet.cs.kuleuven.be/software/linddun/catalog.php

[Wuyts and Joosen, 2015] Kim Wuyts and Wouter Joosen, LINDDUN privacy threat modeling: a tutorial, *Technical Report (CW Reports)*, **C685**, Department of Computer Science, KU Leuven, 2015.

[Wuyts et al., 2015] Kim Wuyts, Riccardo Scandariato, and Wouter Joosen. LINDDUN evaluation - experimental material, available as:
https://sites.google.com/site/linddunstudy/

[European, 1995] European Parliament. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data., 1995.

[Pfitzmann and Hansen, 2010] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. TU Dresden and ULD Kiel, Report V-2010-0.33, April 2010.

[Roe et al., 1997] Michael Roe. Cryptography and Evidence. PhD thesis, University of Cambridge, Clare College, 1997.

[McCallister et al., 2009] Erika McCallister, Tim Grance, and Karen Kent. Guide to protecting the confidentiality of personally identifiable information (PII), National Institute of Standards and Technology, Report, 2009.

[Endsley, 1995] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **37**(1), 1995, 32–64.

[Dourish and Bellotti, 1992] Paul Dourish and Victoria Bellotti, Awareness and coordination in shared workspaces, *In Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work*, **CSCW92**, 1992, 107–114.

[Sohlenkamp, 1998] Markus Sohlenkamp. Supporting Group Awareness in Multi-User Environments Through Perceptualization. PhD thesis, Universität Paderborn, Germany, 1998.

[Craig Larman, 2005] Craig Larman, Applying UML and Patterns, *Pearson Education*, 2005.

[OWASP, 2015] The Open Web Application Software Project. Threat Risk Modeling, 2015, available as:
https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology#Approach

[Wuyts et al., 2015] Kim Wuyts, Privacy threat trees catalog. Available as: https://people.cs.kuleuven.be/~kim.wuyts/LINDDUN/, 2015.

[Kalloniatis et al., 2008] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. Addressing privacy requirements in system design: the PriS method, *Requirements Engineering*, **13**(3), 2008, 241–255.

[Wuyts et al., 2009] Kim Wuyts, Riccardo Scandariato, Bart De Decker, and Wouter Joosen. Linking privacy solutions to developer goals, *In Proceedings of the Fourth International Conference on Availability, Security and Reliability*, IEEE Computer Society, March, 2009, 847–852.

Appendix 1

Data 1: Privacy threats produced by LINDDUN

Threat 01. Login threat.

Summary	Users' login information (account ID, password, profile…) is linkable to other logins, or too specific which can easily be an identity. The account will be linkable to other services.
Primary misactor	User.
Basic flow	Bf1: Users use fixed information as id (or password, profile…) to login. Bf2: Potential attackers find similar ID or related information, then get more information which should not be open. Consequence: Potential attackers get more misactors' information than expected.
*Leaf node(s)**	L_e2, L_e4, L_e5
*Root node(s)**	Linkability of entity
*DFD element(s)**	Normal user

Threat 02. Untrustworthy information receiver threat..

Summary	Information receivers (service provider), or a part of users are not trustable.
Primary misactor	User, information receiver.
Basic flow	Bf1: Users provide information. Bf2: The information is disclosed by misactor. Consequence: Potential attackers get users' information through misactor.
*Leaf node(s)**	L_e3, L_e6, I_e3, I_e6
*Root node(s)**	Linkability of entity, identifiability of entity, non-compliance.
*DFD element(s)**	Normal user

Threat 03. Too much information sharing threat.

Summary	Information receivers make bad user friendly privacy default setting. Users share too much information, and information receivers don't minimize these information well, which is harmful to users' privacy.
Primary misactor	User, information receiver, unskilled or intentional insider.
Basic flow	Bf1: Users share too much information. Bf2: Information receiver does not minimize the information well.

	Bf3: Potential attackers get users' privacy information by inference. Consequence: Potential attackers get more misactors' information than expected.
Alter flow	Af1: Users share too much information. Af2: Information receiver does not minimize the information well. Af3: Potential attackers get users' identifiable information by inference. Consequence: Potential attackers get more misactors' information than expected.
*Leaf node(s)**	I_e2, I_e4, I_e5, I_e8, I_e12, I_e13, I_e17, I_e8, U_1, U_3.
*Root node(s)**	Linkability of entity, identifiability of entity, unawareness.
*DFD element(s)**	Normal user

Threat 04. Unclear or unawareness storage of information.

Summary	The RTT information receiver collects user's information. The user cannot check if the information receiver collects data or not.
Primary misactor	User, unskilled or intentional insider.
Basic flow	Bf1: Information receivers collect and store users' information. Bf2: Users cannot check and manage their information well, which increase the possibility of information disclosure. Consequence: Potential attackers get information without user's awareness.
*Leaf node(s)**	U_2, U_5.
*Root node(s)**	Unawareness
*DFD element(s)**	Normal user

Threat 05. Spoofing a user of the RTT system by eavesdropping communication.

Summary	The communication between user and system portal is not protected well. Potential attackers have chance to falsify user credential to spoof a user.
Primary misactor	Skilled outsider (attacker).
Basic flow	Bf1: The misactor makes fake credential to login. Bf2: The misactor gets target user's information. Consequence: Users' information is disclosed.
*Leaf node(s)**	L_df3, I_df3, L_df6, I_df6.
*Root node(s)**	Linkability and identifiability of data flow.
*DFD element(s)**	Df_1, 2, 3, 4.

Threat 06. Spoofing a user of the RTT system by falsifying credentials.

Summary	The communication between user and system portal is not protected well. Potential attackers have chance to eavesdrop credential communication to spoof a user.
Primary misactor	Skilled outsider (attacker).
Basic flow	Bf1: The misactor eavesdrops credential communication to get key information. Bf2: The misactor spoofs a user as an authorized user. Bf3: The misactor gets target user's information. Consequence: Users' information is disclosed.
*Leaf node(s)**	L_df3, I_df3, L_df6, I_df6.
*Root node(s)**	Linkability and identifiability of data flow.
*DFD element(s)**	Df_1, 2, 3, 4.

Threat 07. Credential is disclosed.

Summary	The credential of users or content seller is disclosed.
Primary misactor	Skilled outsider (attacker).
Basic flow	Bf1: The credential of users or content seller is disclosed. Bf2: The misactor gets the credential and are able to access data. Consequence: All data in the RTT system become in danger.
*Leaf node(s)**	ID_df2
*Root node(s)**	Information disclosure of data flow.
*DFD element(s)**	Df_1, 2, 3, 4.

Threat 08. Communication session token is disclosed.

Summary	The communication token of one action in the RTT system is disclosed.
Primary misactor	Skilled outsider (attacker).
Basic flow	Bf1: The communication token of one action in the RTT system is disclosed. Bf2: The misactor gets the communication token and get more data somehow. For example, spoofing a user or content seller. Consequence: Data related to disclosed communication become in dinger.
*Leaf node(s)**	ID_df2
*Root node(s)**	Information disclosure of data flow.
*DFD element(s)**	All data flows

Threat 09. Information disclosure of internal process.

Summary	The misactor is able to access an internal process, which should be out of his/her range.
Primary misactor	Unskilled or intentional insider.
Basic flow	Bf1: The misactor makes good use of privilege to get access to target process, such as location process. Bf2: The misactor gets data. Consequence: Users' data becomes in danger.
*Leaf node(s)**	ID_p1
*Root node(s)**	Information disclosure of process.
*DFD element(s)**	All processes

Threat 10. Communication content is disclosed.

Summary	The communication content of one action in the RTT system is disclosed.
Primary misactor	Skilled outsider (attacker).
Basic flow	Bf1: The communication token of one action in the RTT system is disclosed. Bf2: The misactor gets the communication content. Consequence: Users' data becomes in danger.
*Leaf node(s)**	ID_df1
*Root node(s)**	Information disclosure of data flow.
*DFD element(s)**	All data flows

Threat 11. Communication with third party is not protected.

Summary	Data is transmitted through an unprotected way to third party. The session token and communication content are both not protected well.
Primary misactor	Skilled outsider (attacker).
Basic flow	Bf1: Users use third-party services through the RTT system. Or users use the RTT service with third-party account information support. Bf2: The communication token or content is disclosed. Bf3: The misactor gets the communication token, or get data directly. Consequence: Users' data becomes in danger.
*Leaf node(s)**	ID_df1, L_df5, I_df5.
*Root node(s)**	Information disclosure of data flow.
*DFD element(s)**	Df_27, 28, 31, 32, 33, 34

Threat 12. Third party is untrustworthy.

Summary	Third party information receivers are not trustable.
Primary misactor	Information receiver, unskilled or intentional insider.
Basic flow	Bf1: Users use third-party services through the RTT system. Or users use the RTT service with third-party account information support. Bf2: The misactor discloses users' information. Or the misactor does not protect users' information well. Consequence: Users' data is disclosed.
*Leaf node(s)**	ID_ds
*Root node(s)**	Linkability, identifiability, information disclosure of data store.
*DFD element(s)**	Bank database, google map database, other third party databases.

Threat 13. Database is untrustworthy.

Summary	The RTT database, namely Mongo database is untrustworthy, or, not protected well.
Primary misactor	Unskilled or intentional insider.
Basic flow	Bf1: User information is stored in the RTT database. Bf2: Potential attackers steal data from the RTT database. Consequence: Users' data is disclosed.
*Leaf node(s)**	ID_ds
*Root node(s)**	Linkability, identifiability, information disclosure of data store.
*DFD element(s)**	Mongo database

Threat 14. Incorrect or insufficient privacy policies.

Summary	Privacy policy for RTT is incorrect or insufficient, which makes users unconsent on collection and usage of user data.
Primary misactor	Unskilled or intentional insider.
Basic flow	Bf1: Users accept incorrect or insufficient privacy policies. Or users do not read privacy policies. Bf2: Information receivers collect and store users' information for untrustworthy purposes. Consequence: Users' information is disclosed to the misactor.
*Leaf node(s)**	NC_2, NC_3, NC_4.
*Root node(s)**	Non-compliance
*DFD element(s)**	All elements except entity.

Threat 15. Non-compliance insider actions

Summary	Skilled insiders intentionally access data and disclose data.
Primary misactor	Intentional insider.
Basic flow	Bf1: Skilled insiders access users' data. Bf2: Skilled insiders share data with potential attackers. Consequence: Users' information is disclosed to the misactor.
*Leaf node(s)**	NC_2, NC_3, NC_4.
*Root node(s)**	Non-compliance
*DFD element(s)**	All elements except entity.

Appendix 2

LINDDUN workshop interview questions:

This interview is for LINDDUN workshop participants. Please answer the questions after you finish the workshop issues.

1. How many projects have you involved totally? Do you have experience on requirements engineering in previous projects?
2. Value the output of your work: Data Flow Diagram, from no help (0) to great help (5).
3. Value the output of your work: Privacy threats list, from no help (0) to great help (5).
4. Which part is most difficult in the workshop? Which part is most useful in the workshop?
5. What support do you mostly want in the workshop? When do you feel like you need help?
6. What method or tool did you use in the workshop? How does this help you?
7. Anything else?